(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 February 2006 (09.02.2006)

PCT

(10) International Publication Number
WO 2006/015365 A1

(51) International Patent Classification:
C07K 14/00 (2006.01) *C12N 15/693* (2006.01)
C12N 5/10 (2006.01) *G01N 33/53* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2005/027579

(22) International Filing Date: 1 August 2005 (01.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/592,592 30 July 2004 (30.07.2004) US

(71) Applicant *(for all designated States except US)*: **MOUNT SINAI SCHOOL OF MEDICINE** [US/US]; One Gustave L. Levy Place, New York, NY 10029 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **IOANNOU, Yiannis** [US/US]; 306 E. 96th Street, Apt. #14E, New York, NY 10128 (US). **DAVIES, Joanna, P.** [US/US]; 2237 36th Street, Long Island City, NY 11105 (US).

(74) Agents: **LUDWIG, S., Peter** et al.; Darby & Darby P.C., P.O. Box 5257, New York, NY 10150-5257 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2006/015365 A1

(54) Title: NPC1L1 AND NPC1L1 INHIBITORS AND METHODS OF USE THEREOF

(57) Abstract: The present invention provides a novel gene, designated herein as "NPC1L1", that is associated with lipid or glucose metabolism. The invention further provides the use of the NPC1L1 gene and its corresponding protein to diagnose a lipid condition in a cell or tissue and to screen for novel therapeutic compounds useful for treating lipid disorders and other NPC1L1-associated or mediated diseases or disorders. The invention further provides specific inhibitors of NPC1L1.

NPC1L1 AND NPC1L1 INHIBITORS AND METHODS OF USE THEREOF

RELATED APPLICATIONS

The present application claims priority to provisional application Serial No. 60/592,592, filed on July 30, 2004, the contents of which are expressly incorporated by reference herein.

FIELD OF INVENTION

The present invention relates to the identification of a Niemann-Pick C1 Like 1 (NPC1L1) gene. The present invention further includes NPC1L1 nucleic acids and polypeptides, as well as transgenic animals with disrupted NPC1L1 function. In addition, the present invention relates to methods of use for NPC1L1 molecules, including drug screening, diagnostics, and treatment of disorders relating to aberrant lipid and glucose metabolism.

BACKGROUND OF THE INVENTION

Lipid Metabolism and Hyperlipidemia

Diets high in lipids, such as fat and cholesterol, are important factors in the development of many human diseases, including obesity, diabetes mellitus, atherosclerosis, and coronary artery disease. In addition, aberrant regulation of lipids can contribute to many other conditions, such as arthritis, cancer, hypertension, and vascular disorders. Modulating the biochemical and molecular mechanisms of lipid metabolism is therefore a crucial goal of contemporary research and medicine.

The control of lipid metabolism is highly complex, reflecting a delicate balance between the processes of ingestion, synthesis, and mobilization. The mechanisms underlying cholesterol control, for example, include absorption of dietary

cholesterol in the intestine; *de novo* production of cholesterol in the liver; secretion of cholesterol into the blood and lymph via lipoprotein carriers, and transport of cholesterol-lipoproteins from the serum to target tissues for use and elimination. Each of these steps represents a potential point for regulation as well as potential target for medical intervention.

In addition, chemical modifications of lipids play a key role in regulating metabolism. One key step is the addition of ester groups to cholesterol in the endoplasm reticulum, a modification that renders cholesterol more hydrophobic and competent for assembly into lipoprotein complexes. Lipoprotein complexes are essential for the transport of lipids to tissues; free lipids are virtually undetectable in the blood. There are least five distinct families of lipoproteins, each distinguished by their density as well as functional role in lipid metabolism.

Cholesterol esters are not just critical in intestinal absorption of cholesterol and its subsequent deposition into lipoprotein carriers. They are also the major component of atherosclerotic plaques, which underlie vascular disorders such as coronary artery disease--the leading cause of death in industrialized nations. Accordingly, the aberrant regulation of cholesterol metabolism can lead to elevated levels of serum cholesterol and promote cardiovascular disease.

While the pathways underlying *de novo* synthesis and breakdown of cholesterol are well understood, the specific mechanisms that mediate cholesterol transport across the intestinal epithelium remains unclear. Finding new ways to block the absorption of cholesterol may lower serum cholesterol and have significant clinical implications for conditions such as diet-induced obesity, diabetes, and cardiovascular disease. There is a need in the art for further investigations of lipid metabolism, especially with respect to cholesterol absorption.

Niemann Pick C1

The human Niemann-Pick C1 gene (NPC1) encodes a transmembrane transporter that is defective in the rare cholesterol storage disease, Niemann-Pick C1. NPC1 localizes to late endosomes and plays a pivotal role in intracellular transport of cholesterol and other lipids. Cells lacking NPC1 have a number of distinct trafficking

defects: (i) unesterified cholesterol derived from low-density lipoproteins (LDLs) accumulates in lysosomes; (ii) cholesterol accumulates in the trans-golgi network; and (iii) cholesterol transport to and from the plasma membrane is delayed.

The present invention provides a novel Niemann-Pick C1 Like 1 (NPC1L1) gene that is also involved in lipid metabolism.

SUMMARY OF THE INVENTION

The present invention provides an isolated nucleic acid that comprises a nucleotide sequence encoding a non-human NPC1L1 polypeptide, and fragments thereof. In one embodiment, the isolated genomic nucleic acid comprises a nucleotide sequence set forth SEQ ID NO:1.

In another embodiment, the nucleic acid comprises a nucleotide sequence set forth SEQ ID NO:2.

The present invention provides an isolated NPC1L1 nucleic acid which encodes a polypeptide having an amino acid sequence set forth in SEQ ID NO:3.

The present invention also provides NPC1L1 polypeptides encoded by the NPC1L1 nucleic acid sequences described above. In one embodiment, the NPC1L1 polypeptide is a non-human NPC1L1 polypeptide. In a specific embodiment, embodiment, the NPC1L1 polypeptide has the amino acid sequence set forth in SEQ ID NO: 3.

In addition, the present invention encompasses isolated nucleic acids with mutations in NPC1L1 coding sequences, and which encode NPC1L1 polypeptides having altered amino acid sequences.

The invention also provides recombinant vectors and host cells comprising the NPC1L1 nucleic acid molecules, as well as methods for producing an NPC1L1 polypeptide using such host cells. In one embodiment, the host cells are bacterial or eukaryotic cells engineered for studies of NPC1L1 function.

The invention further provides non-human transgenic animals comprising such a recombinant vector. In one embodiment, the animal is a mouse.

The invention also provides an oligonucleotide, such as a primer or probe, wherein the oligonucleotide has a sequence identical to a contiguous nucleotide sequence in the NPC1L1 nucleotide sequence, e.g., SEQ ID NO:2. The oligonucletide has a length at least 10 bases, preferably at least 20 bases, and more preferably at least 30 bases.

The invention further provides antibodies that bind specifically to an NPC1L1 protein having an amino acid sequence shown in SEQ ID NO:3, or fragments thereof.

The present invention includes methods of screening to identify an antagonist or agonist of a NPC1L1 nucleic acid or polypeptide. Such agonists/antagonists are thus designated candidate compounds for the treatment (*e.g.*, therapeutic and prophylactic) of NPC1L1-mediated disorders, such as hyperlipidemia, and other diseases and disorders associated with or mediated by NPC1L1, including, but not limited to, body weight disorders such as obesity, diabetes, *e.g.*, type II diabetes, cardiovascular disease, including, for example, ischemia, congestive heart failure, and atherosclerosis, and stroke. NPC1L1-mediated disorders include those disorders which are mediated by the expression or activity of NPC1L1, including plasma membrane uptake and transport of various lipids, including cholesterol and sphingolipids.

In one embodiment, the NPC1L1 antagonist is selected from the group consisting of a small molecule, an anti-NPC1L1 antibody, an NPC1L1 antisense nucleic acid, an NPC1L1 ribozyme, an NPC1L1 triple-helix, or an NPC1L1 inhibitory RNA. In another embodiment, the NPC1L1 antagonist inhibits transcription of NPC1L1 by targeting an NPC1L1 promoter transcription factor. In this embodiment the specific agonist or antagonist is identified by its ability to downregulate the expression of a reporter gene (such as luciferase or green fluorescence protein) driven by the promoter for NPC1L1. In another embodiment, the inhibitor is selected from the group consisting on: 4-phenyl-4-piperidinecarbonitrile hydrochloride, 1-butyl-N-(2,6-dimethylphenyl)-2 piperidinecarboxamide, 1-(1-naphthylmethyl)piperazine, 3{1-[(2-methylphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3-[(4-methoxyphenyl)amino]-2-methyl-2-

cyclopenten-1-one, 3-[(2-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, and N-(4-acetylphenyl)-2-thiophenecarboxamide.

The invention further provides a mammal, preferably a mouse, comprising a homozygous or heterozygous disruption of endogenous NPC1L1, wherein the mouse produces less functional NPC1L1 polypeptide or does not produce any functional NPC1L1 polypeptide.

The invention further describes transgenic mammal, preferably a mouse, in which the mouse NPC1L1 genomic gene or cDNA is into the mouse genome in multiple copies, which is a model for hyperlipidemia. In one embodiment, the hyperlipidemia is hypercholesterolemia.

The present invention also provides a method of inhibiting the cellular uptake of a lipid by inhibiting the expression or activity of an NPC1L1 nucleic acid or polypeptide.

Further provided is a method of treating hyperlipidemia or other diseases and disorders associated with or mediated by NPC1L1, including, but not limited to, obesity, diabetes, *e.g.*, type II diabetes, cardiovascular disease, or stroke in a subject in need thereof by administering to the subject a therapeutically effective amount of an agent which inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

In one embodiment, the NPC1L1 nucleic acid or polypeptide which is inhibited is that set forth in SEQ ID NOs: 2 and 3, respectively.

In another embodiment, the hyperlipidemia is hypercholesterolemia.

The present invention further provides a method of decreasing the plasma glucose by administering a therapeutically effective amount of an agent which inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

In one embodiment, the NPC1L1 nucleic acid or polypeptide which is inhibited is that set forth in SEQ ID NOs: 2 and 3, respectively.

In another embodiment, the hyperlipidemia is dietary hypercholesterolemia.

The present invention also provides a method for identifying a test compound that binds to and modulates the activity of an NPC1L1 polypeptide, which compound

is therefore a candidate compound for the treatment of hyperlipidemia, obesity, diabetes, *e.g.*, type II diabetes, cardiovascular disease, or stroke.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1A-1E. Figure 1 demonstrates the subcellular localization of murine NPC1L1 by immunofluorescence. Figure 1a shows localization in human NT2 cells. Figure 1b shows localization of tagged NPC1L1 in transfected COS-7 cells. Figure 1c shows localization in Caco-2 cells transiently transfected with an NPC1L1 fusion protein. Figure 1d depicts the lack of localization of NPC1L1 on the plasma membrane. Figure 1e demonstrates the effect of NPC1L1 on fatty acid transport in bacterial cells.

Figures 2A-2F. Figure 2 shows the tissue distribution of human and mouse NPC1L1 in various tissues in human (Fig. 2a and 2b) and mouse (Fig. 2c) tissues using quantitative real time PCR (Fig. 2d and 2e). Figure 2f demonstrates reduced activation of reporter genes in cells from NPC1L1-deficient mice (L1) compared with control mice (WT), under the expression of three response elements: ABCA1-RFP (Fig. 2f(1-4)); DR4-RFP (Fig. 2f(5-8)); and SRE-GFP (Fig. 2f(9-12)).

Figures 3A-3E. Figure 3 demonstrates impaired uptake of multiple lipids (i.e., oleic acid, cholesterol) in mouse cells from NPC1L1 deficient mice using radioactively labeled lipids (Fig. 3a-b), fluorescently-tagged lipids complexed with cyclodextrin (Fig. 3c) or BSA (Fig. 3d). Figure 3e demonstrates expression of a caveolin-mYFP fusion in mouse wild-type or NPC1L1 null cells.

Figure 4. Figure 4 demonstrates resistance to hypercholesterolemia in NPC1L1 null mice subjected to a high cholesterol diet. Figure 4 shows plasma assays for glucose, triglycerides, total cholesterol and HDL-cholesterol after 14 weeks.

Figure 5. Figure 5 demonstrates the AcrAB-TolC complex in *E. coli* and the homologous MexCD-OprJ complex from *Pseudomonas aeruginosa*.

Figure 6. Immunofluorescence of lysosomal cholesterol of normal human fibroblasts treated (6B) or untreated (6A) with NPC1 inhibitor 4-butyryl-4-phenylpiperidine.

Figure 7. Immunofluorescence of lysosomal cholesterol of normal human fibroblasts treated with weaker NPC1 inhibitor 4-cyano-4-phenylpiperidine (7A), or 4-methylpiperidine (7B).

Figure 8 is a graph illustrating that inhibitors 4-Phenyl-4-piperidinecarbonitrile Hydrochloride (#1), (1-Butyl-N(2,6-diemethylphenyl)2 piperidine carboxamide) #7, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione and gave a positive signal compared to control (none). Note that Ezetamibe did not inhibit NPC1L1 in this assay.

Figures 9A-9B. Figure 9A is a graph depicting body weights of mice fed a high fat diet for 0-245 days (Mouse set 1). **Figure 9B** is a graph depicting body weights of mice fed a high fat diet for 0-95 days (mouse set 2).

Figure 10 is a graph depicting results of a glucose tolerance test on mice fed with regular chow (mouse set 1).

Figures 11A-11B. Figure 11A is a graph depicting results of a glucose tolerance test on mice fed a high fat diet for 102 days (mouse set 1). **Figure 11B** is a graph depicting results of a glucose tolerance test on mice fed a high fat diet for 262 days (mouse set 1).

Figures 12A-12B. Figure 12A is a graph depicting results of an insulin tolerance test in mice fed a high fat diet for 105 days (mouse set 2). **Figure 12B** is a graph depicting results of an insulin tolerance test in mice fed a high fat diet for 252 days (mouse set 1).

Figures 13A-13B. Figure 13A is a graph depicting insulin measurements in mice fed a high fat diet for 72 days (mouse set 2). **Figure 13B** is a graph depicting insulin measurements in mice fed a high fat diet for 220 days (mouse set 1).

Figures 14A-14B are graphs depicting plasma lipoprotein profiles in mice at 120 days (Figure 14A) and 268 days (Figure 14B) of high fat diet.

Figure 15 is a graph depicting results of real-time PCR of NPC1L1 in mouse tissue and 3T3L1 cell line.

Figure 16 is a graph depicting results of real-time PCR of NPC1L1 in mouse white and brown adipose tissue.

Figure 17 is a graph depicting results of real-time PCR of NPC1L1 in human liver and adipose tissue.

Figure 18 is a table illustrating weight gain and food intake over 210 days for NPC1L1 knockout mice fed a high fat diet as compared to wild type mice fed a high fat diet.

DETAILED DESCRIPTION OF THE INVENTION

The Niemann Pick C1-like gene and gene product (NPC1L1; also known as NPC3; Genbank Accession No. AF192522; Davies et al., (2000) Genomics 65(2): 137-145 and Ioannou et al., (2000) Mol. Genet. Metab. 71(1-2): 175-181 was first isolated in humans, based on its 42% amino acid identity and 51% amino acid similarity to human NPC1 (Genbank Accession No. AF002020).

The present invention is based on methods of using NPC1L1 molecules including screening assays for identifying modulators of NPC1L1, inhibitors of NPC1L1 including small molecule compounds, antibodies, and siRNA molecules, NPC1L1 knock-out animals and transgenic animals, as well as therapeutic methods for the treatment of NPC1L1 mediated disease and disorders including, but not limited to, lipid disorders such as hyperlipidemia, and obesity, diabetes, and cardiovascular disease using modulators, *e.g.*, inhibitors of NPC1L1. Methods for treating disorders associated with decreased NPC1L1, *e.g.*, anorexia, cachexia, and wasting, using agonists of NPC1L1 are also included in the invention. The present invention also includes diagnostic methods using NPC1L1.

Definitions

The term "subject" as used herein refers to a mammal (*e.g.*, a rodent such as a mouse or a rat, a pig, a primate, or companion animal (*e.g.*, dog or cat, etc.). In particular, the term refers to humans.

The terms "array" and "microarray" are used interchangeably and refer generally to any ordered arrangement (*e.g.*, on a surface or substrate) of different

molecules, referred to herein as "probes." Each different probe of an array is capable of specifically recognizing and/or binding to a particular molecule, which is referred to herein as its "target," in the context of arrays. Examples of typical target molecules that can be detected using microarrays include mRNA transcripts, cDNA molecules, cRNA molecules, and proteins. As disclosed in the Examples section below, at least one target detectable by the Affymetrix GeneChip® microarray used as described herein is a NPC1L1-encoding nucleic acid (such as an mRNA transcript, or a corresponding cDNA or cRNA molecule).

An "antisense" nucleic acid molecule or oligonucleotide is a single stranded nucleic acid molecule, which may be DNA, RNA, a DNA-RNA chimera, or a derivative thereof, which, upon hybridizing under physiological conditions with complementary bases in an RNA or DNA molecule of interest, inhibits the expression of the corresponding gene by inhibiting, *e.g.*, mRNA transcription, mRNA splicing, mRNA transport, or mRNA translation or by decreasing mRNA stability. As presently used, "antisense" broadly includes RNA-RNA interactions, RNA-DNA interactions, and RNase-H mediated arrest. Antisense nucleic acid molecules can be encoded by a recombinant gene for expression in a cell (see, *e.g.*, U.S. Patents No. 5,814,500 and 5,811,234), or alternatively they can be prepared synthetically (see, *e.g.*, U.S. Patent No. 5,780,607). According to the present invention, the role of NPC1L1 in regulation of conditions associated with hyperlipidemia may be identified, modulated and studied using antisense nucleic acids derived on the basis of NPC1L1-encoding nucleic acid molecules of the invention.

The term "ribozyme" is used to refer to a catalytic RNA molecule capable of cleaving RNA substrates. Ribozyme specificity is dependent on complementary RNA-RNA interactions (for a review, see Cech and Bass, *Annu. Rev. Biochem.* 1986; 55: 599-629). Two types of ribozymes, hammerhead and hairpin, have been described. Each has a structurally distinct catalytic center. The present invention contemplates the use of ribozymes designed on the basis of the NPC1L1-encoding nucleic acid molecules of the invention to induce catalytic cleavage of the corresponding mRNA, thereby inhibiting expression of the NPC1L1 gene. Ribozyme

technology is described further in *Intracellular Ribozyme Applications: Principals and Protocols*, Rossi and Couture ed., Horizon Scientific Press, 1999.

The term "RNA interference" or "RNAi" refers to the ability of double stranded RNA (dsRNA) to suppress the expression of a specific gene of interest in a homology-dependent manner. It is currently believed that RNA interference acts post-transcriptionally by targeting mRNA molecules for degradation. RNA interference commonly involves the use of dsRNAs that are greater than 500 bp; however, it can also be mediated through small interfering RNAs (siRNAs) or small hairpin RNAs (shRNAs), which can be 10 or more nucleotides in length and are typically 18 or more nucleotides in length. For reviews, see Bosner and Labouesse, *Nature Cell Biol.* 2000; 2: E31-E36 and Sharp and Zamore, *Science* 2000; 287: 2431-2433.

The term "nucleic acid hybridization" refers to anti-parallel hydrogen bonding between two single-stranded nucleic acids, in which A pairs with T (or U if an RNA nucleic acid) and C pairs with G. Nucleic acid molecules are "hybridizable" to each other when at least one strand of one nucleic acid molecule can form hydrogen bonds with the complementary bases of another nucleic acid molecule under defined stringency conditions. Stringency of hybridization is determined, *e.g.*, by (i) the temperature at which hybridization and/or washing is performed, and (ii) the ionic strength and (iii) concentration of denaturants such as formamide of the hybridization and washing solutions, as well as other parameters. Hybridization requires that the two strands contain substantially complementary sequences. Depending on the stringency of hybridization, however, some degree of mismatches may be tolerated. Under "low stringency" conditions, a greater percentage of mismatches are tolerable (*i.e.*, will not prevent formation of an anti-parallel hybrid). See *Molecular Biology of the Cell*, Alberts *et al.*, 3rd ed., New York and London: Garland Publ., 1994, Ch. 7.

Typically, hybridization of two strands at high stringency requires that the sequences exhibit a high degree of complementarity over an extended portion of their length. Examples of high stringency conditions include: hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% SDS, 1 mM EDTA at 65°C, followed by washing in 0.1x SSC/0.1% SDS (where 1x SSC is 0.15 M NaCl, 0.15 M Na citrate) at 68°C or for

oligonucleotide molecules washing in 6xSSC/0.5% sodium pyrophosphate at about 37°C (for 14 nucleotide-long oligos), at about 48°C (for about 17 nucleotide-long oligos), at about 55°C (for 20 nucleotide-long oligos), and at about 60°C (for 23 nucleotide-long oligos)).

Conditions of intermediate or moderate stringency (such as, for example, an aqueous solution of 2×SSC at 65°C; alternatively, for example, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% SDS, 1 mM EDTA at 65°C, and washing in 0.2 x SSC/0.1% SDS at 42°C) and low stringency (such as, for example, an aqueous solution of 2×SSC at 55°C), require correspondingly less overall complementarity for hybridization to occur between two sequences. Specific temperature and salt conditions for any given stringency hybridization reaction depend on the concentration of the target DNA and length and base composition of the probe, and are normally determined empirically in preliminary experiments, which are routine (see Southern, *J. Mol. Biol.* 1975; 98: 503; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 2, ch. 9.50, CSH Laboratory Press, 1989; Ausubel *et al.* (eds.), 1989, *Current Protocols in Molecular Biology*, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York, at p. 2.10.3).

As used herein, the term "standard hybridization conditions" refers to hybridization conditions that allow hybridization of two nucleotide molecules having at least 75% sequence identity. According to a specific embodiment, hybridization conditions of higher stringency may be used to allow hybridization of only sequences having at least 80% sequence identity, at least 90% sequence identity, at least 95% sequence identity, or at least 99% sequence identity.

Nucleic acid molecules that "hybridize" to any of the NPC1L1-encoding nucleic acids of the present invention may be of any length. In one embodiment, such nucleic acid molecules are at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, and at least 70 nucleotides in length. In another embodiment, nucleic acid molecules that hybridize are of about the same length as the particular NPC1L1-encoding nucleic acid.

The term "homologous" as used in the art commonly refers to the relationship between nucleic acid molecules or proteins that possess a "common evolutionary origin," including nucleic acid molecules or proteins within superfamilies (*e.g.*, the immunoglobulin superfamily) and nucleic acid molecules or proteins from different species (Reeck *et al.*, *Cell* 1987; 50: 667). Such nucleic acid molecules or proteins have sequence homology, as reflected by their sequence similarity, whether in terms of substantial percent similarity or the presence of specific residues or motifs at conserved positions.

The terms "percent (%) sequence similarity", "percent (%) sequence identity", and the like, generally refer to the degree of identity or correspondence between different nucleotide sequences of nucleic acid molecules or amino acid sequences of proteins that may or may not share a common evolutionary origin (see Reeck *et al.*, *supra*). Sequence identity can be determined using any of a number of publicly available sequence comparison algorithms, such as BLAST, FASTA, DNA Strider, GCG (Genetics Computer Group, Program Manual for the GCG Package, Version 7, Madison, Wisconsin), etc.

In addition to the NPC1L1 nucleic acid sequences and NPC1L1 polyleptides (as shown in, e.g., SEQ ID NOS: 2 and 3, respectively), the present invention further provides polynucleotide molecules comprising nucleotide sequences having certain percentage sequence identities to any of the aforementioned sequences. Such sequences preferably hybridize under conditions of moderate or high stringency as described above, and may include species orthologs.

As used herein, the term "orthologs" refers to genes in different species that apparently evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function through the course of evolution. Identification of orthologs can provide reliable prediction of gene function in newly sequenced genomes. Sequence comparison algorithms that can be used to identify orthologs include without limitation BLAST, FASTA, DNA Strider, and the GCG pileup program. Orthologs often have high sequence similarity.

The present invention encompasses all non-human orthologs of NPC1L1. In addition to the mouse ortholog, particularly useful NPC1L1 orthologs of the present invention are rat, monkey, porcine, canine (dog), and guinea pig orthologs.

As used herein, the term "isolated" means that the material being referred to has been removed from the environment in which it is naturally found, and is characterized to a sufficient degree to establish that it is present in a particular sample. Such characterization can be achieved by any standard technique, such as, *e.g.*, sequencing, hybridization, immunoassay, functional assay, expression, size determination, or the like. Thus, a biological material can be "isolated" if it is free of cellular components, *i.e.*, components of the cells in which the material is found or produced in nature. For nucleic acid molecules, an isolated nucleic acid molecule or isolated polynucleotide molecule, or an isolated oligonucleotide, can be a PCR product, an mRNA transcript, a cDNA molecule, or a restriction fragment. A nucleic acid molecule excised from the chromosome that it is naturally a part of is considered to be isolated. Such a nucleic acid molecule may or may not remain joined to regulatory, or non-regulatory, or non-coding regions, or to other regions located upstream or downstream of the gene when found in the chromosome. Nucleic acid molecules that have been spliced into vectors such as plasmids, cosmids, artificial chromosomes, phages and the like are considered isolated. In a particular embodiment, a NPC1L1-encoding nucleic acid spliced into a recombinant vector, and/or transformed into a host cell, is considered to be "isolated".

Isolated nucleic acid molecules and isolated polynucleotide molecules of the present invention do not encompass uncharacterized clones in man-made genomic or cDNA libraries.

A protein that is associated with other proteins and/or nucleic acids with which it is associated in an intact cell, or with cellular membranes if it is a membrane-associated protein, is considered isolated if it has otherwise been removed from the environment in which it is naturally found and is characterized to a sufficient degree to establish that it is present in a particular sample. A protein expressed from a recombinant vector in a host cell, particularly in a cell in which the protein is not naturally expressed, is also regarded as isolated.

An isolated organelle, cell, or tissue is one that has been removed from the anatomical site (cell, tissue or organism) in which it is found in the source organism.

An isolated material may or may not be "purified". The term "purified" as used herein refers to a material (*e.g.*, a nucleic acid molecule or a protein) that has been isolated under conditions that detectably reduce or eliminate the presence of other contaminating materials. Contaminants may or may not include native materials from which the purified material has been obtained. A purified material preferably contains less than about 90%, less than about 75%, less than about 50%, less than about 25%, less than about 10%, less than about 5%, or less than about 2% by weight of other components with which it was originally associated.

Methods for purification are well-known in the art. For example, nucleic acids or polynucleotide molecules can be purified by precipitation, chromatography (including preparative solid phase chromatography, oligonucleotide hybridization, and triple helix chromatography), ultracentrifugation, and other means. Polypeptides can be purified by various methods including, without limitation, preparative disc-gel electrophoresis, isoelectric focusing, HPLC, reverse-phase HPLC, gel filtration, affinity chromatography, ion exchange and partition chromatography, precipitation and salting-out chromatography, extraction, and counter-current distribution. Cells can be purified by various techniques, including centrifugation, matrix separation (*e.g.*, nylon wool separation), panning and other immunoselection techniques, depletion (*e.g.*, complement depletion of contaminating cells), and cell sorting (*e.g.*, fluorescence activated cell sorting (FACS)). Other purification methods are possible. The term "substantially pure" indicates the highest degree of purity that can be achieved using conventional purification techniques currently known in the art. In the context of analytical testing of the material, "substantially free" means that contaminants, if present, are below the limits of detection using current techniques, or are detected at levels that are low enough to be acceptable for use in the relevant art, for example, no more than about 2-5% (w/w). Accordingly, with respect to the purified material, the term "substantially pure" or "substantially free" means that the purified material being referred to is present in a composition where it represents 95% (w/w) or more of the weight of that composition. Purity can be evaluated by

chromatography, gel electrophoresis, immunoassay, composition analysis, biological assay, or any other appropriate method known in the art.

The term "about" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, *i.e.*, the limitations of the measurement system. For example, "about" can mean within an acceptable standard deviation, per the practice in the art. Alternatively, "about" can mean a range of up to ±20%, preferably up to ±10%, more preferably up to ±5%, and more preferably still up to ±1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated, the term "about" is implicit and in this context means within an acceptable error range for the particular value.

The term "degenerate variants" of a polynucleotide sequence are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position.

The term "modulator" refers to a compound that differentially affects the expression or activity of a gene or gene product (*e.g.*, nucleic acid molecule or protein), for example, in response to a stimulus that normally activates or represses the expression or activity of that gene or gene product when compared to the expression or activity of the gene or gene product not contacted with the stimulus. In one embodiment, the gene or gene product the expression or activity of which is being modulated includes a gene, cDNA molecule or mRNA transcript that encodes a mammalian NPC1L1 protein such as, *e.g.*, a rat, mouse, companion animal, or human NPC1L1 protein.

An "antagonist" is one type of modulator, and includes an agent that reduces expression or activity, or inhibits expression or activity, of an NPC1L1 nucleic acid or polypeptide. Examples of antagonists of the NPC1L1-encoding nucleic acids of the present invention include without limitation small molecules, anti-NPC1L1 antibodies, antisense nucleic acids, ribozymes, and RNAi oligonucleotides, and

molecule that target NPC1L1 promoter transcription factors. Specific NPC1L1 antagonists are set forth herein.

An "agonist" is another modulator that is defined as an agent that interacts with (*e.g.*, binds to) a nucleic acid molecule or protein, and promotes, enhances, stimulates or potentiates the biological expression or activity of the nucleic acid molecule or protein. The term "partial agonist" is used to refer to an agonist which interacts with a nucleic acid molecule or protein, but promotes only partial function of the nucleic acid molecule or protein. A partial agonist may also inhibit certain functions of the nucleic acid molecule or protein with which it interacts. An "antagonist" interacts with (*e.g.*, binds to) and inhibits or reduces the biological expression or function of the nucleic acid molecule or protein.

A "test compound" is a molecule that can be tested for its ability to act as a modulator of a gene or gene product. Test compounds can be selected, without limitation, from small inorganic and organic molecules (*i.e.*, those molecules of less than about 2 kD, and more preferably less than about 1 kD in molecular weight), polypeptides (including native ligands, antibodies, antibody fragments, and other immunospecific molecules), oligonucleotides, polynucleotide molecules, and derivatives thereof. In various embodiments of the present invention, a test compound is tested for its ability to modulate the expression of a mammalian NPC1L1-encoding nucleic acid or NPC1L1 protein or to bind to a mammalian NPC1L1 protein. A compound that modulates a nucleic acid or protein of interest is designated herein as a "candidate compound" or "lead compound" suitable for further testing and development. Candidate compounds include, but are not necessarily limited to, the functional categories of agonist and antagonist.

The term "detectable change" as used herein in relation to an expression level of a gene or gene product (*e.g.*, NPC1L1) means any statistically significant change and preferably at least a 1.5-fold change as measured by any available technique such as hybridization or quantitative PCR.

As used herein, the term "specific binding" refers to the ability of one molecule, typically an antibody, polynucleotide, polypeptide, or a small molecule

ligand to contact and associate with another specific molecule, *e.g.*, an NPC1L1 molecule, even in the presence of many other diverse molecules. "Immunospecific binding" refers to the ability of an antibody to specifically bind to (or to be "specifically immunoreactive with") its corresponding antigen.

The term "obesity" or "overweight" is defined as a body mass index (BMI) of 30 kg/ m^2 or more (National Institute of Health, Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults (1998)). However, the present invention is also intended to include a disease, disorder, or condition that is characterized by a body mass index (BMI) of 25 kg/ m^2 or more, 26 kg/m^2 or more, 27 kg/ m^2 or more, 28 kg/ m^2 or more, 29 kg/ m^2 or more, 29.5 kg/ m^2 or more, or 29.9 kg/ m^2 or more, all of which are typically referred to as overweight (National Institute of Health, Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults (1998)). Body weight disorders also include conditions or disorders which are secondary to disorders such as obesity or overweight, *i.e.*, are influenced or caused by a disorder such as obesity or overweight. For example, insulin resistance, diabetes, hypertension, and atherosclerosis can all be influenced or caused by obesity or overweight. Accordingly, such secondary conditions or disorders are additional examples of body weight disorders.

The term "cardiovascular disease" (CVD) is any disease or disorder that affects the cardiovascular system. A cardiovascular disease or disorder includes, but is not limited to atherosclerosis, coronary heart disease or coronary artery disease (CAD), myocardial infarction (MI), ischemia, and peripheral vascular diseases.

"Amplification" of DNA as used herein denotes the use of exponential amplification techniques known in the art such as the polymerase chain reaction (PCR), and non-exponential amplification techniques such as linked linear amplification, that can be used to increase the concentration of a particular DNA sequence present in a mixture of DNA sequences. For a description of PCR, see Saiki *et al.*, *Science* 1988, 239:487 and U.S. Patent No. 4,683,202. For a description of linked linear amplification, see U.S. Patent Nos. 6,335,184 and 6,027,923; Reyes *et al.*, *Clinical Chemistry* 2001; 47: 131-40; and Wu *et al.*, *Genomics* 1989; 4: 560-569.

As used herein, the phrase "sequence-specific oligonucleotides" refers to oligonucleotides that can be used to detect the presence of a specific nucleic acid molecule, or that can be used to amplify a particular segment of a specific nucleic acid molecule for which a template is present. Such oligonucleotides are also referred to as "primers" or "probes." In a specific embodiment, "probe" is also used to refer to an oligonucleotide, for example about 25 nucleotides in length, attached to a solid support for use on "arrays" and "microarrays" described below.

The term "host cell" refers to any cell of any organism that is selected, modified, transformed, grown, used or manipulated in any way so as, *e.g.*, to clone a recombinant vector that has been transformed into that cell, or to express a recombinant protein such as, *e.g.*, a NPC1L1 protein of the present invention. Host cells are useful in screening and other assays, as described below.

As used herein, the terms "transfected cell" and "transformed cell" both refer to a host cell that has been genetically modified to express or over-express a nucleic acid encoding a specific gene product of interest such as, *e.g.*, a NPC1L1 protein or a fragment thereof. Any eukaryotic or prokaryotic cell can be used, although eukaryotic cells are preferred, vertebrate cells are more preferred, and mammalian cells are the most preferred. Transfected or transformed cells are suitable to conduct an assay to screen for compounds that modulate the function of the gene product. A typical "assay method" of the present invention makes use of one or more such cells, *e.g.*, in a microwell plate or some other culture system, to screen for such compounds. The effects of a test compound can be determined on a single cell, or on a membrane fraction prepared from one or more cells, or on a collection of intact cells sufficient to allow measurement of activity.

The term "recombinantly engineered cell" refers to any prokaryotic or eukaryotic cell that has been genetically manipulated to express or over-express a nucleic acid of interest, *e.g.*, a NPC1L1-encoding nucleic acid of the present invention, by any appropriate method, including transfection, transformation or transduction. The term "recombinantly engineered cell" also refers to a cell that has been engineered to activate an endogenous nucleic acid, *e.g.*, the endogenous NPC1L1-encoding gene in a rat, mouse or human cell, which cell would not normally

express that gene product or would express the gene product at only a sub-optimal level.

The terms "vector", "cloning vector" and "expression vector" refer to recombinant constructs including, *e.g.*, plasmids, cosmids, phages, viruses, and the like, with which a nucleic acid molecule (e.g., a NPC1L1-encoding nucleic acid or NPC1L1 siRNA-expressing nucleic acid) can be introduced into a host cell so as to, e.g., clone the vector or express the introduced nucleic acid molecule. Vectors may further comprise selectable markers.

The terms "mutant", "mutated", "mutation", and the like, refer to any detectable change in genetic material, (*e.g.*, NPC1L1 DNA), or any process, mechanism, or result of such a change. Mutations include gene mutations in which the structure (*e.g.*, DNA sequence) of the gene is altered; any DNA or other nucleic acid molecule derived from such a mutation process; and any expression product (*e.g.*, the encoded protein) exhibiting a non-silent modification as a result of the mutation.

As used herein, the term "genetically modified animal" encompasses all animals into which an exogenous genetic material has been introduced and/or whose endogenous genetic material has been manipulated. Examples of genetically modified animals include without limitation transgenic animals, *e.g.*, "knock-in" animals with the endogenous gene substituted with a heterologous gene or an ortholog from another species or a mutated gene, "knockout" animals with the endogenous gene partially or completely inactivated, or transgenic animals expressing a mutated gene or overexpressing a wild-type or mutated gene (*e.g.*, upon targeted or random integration into the genome) and animals containing cells harboring a non-integrated nucleic acid construct (*e.g.*, viral-based vector, antisense oligonucleotide, shRNA, siRNA, ribozyme, etc.), including animals wherein the expression of an endogenous gene has been modulated (*e.g.*, increased or decreased) due to the presence of such construct.

As used herein, a "transgenic animal" is a nonhuman animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of

the cells of the animal include a transgene. Other examples of transgenic animals include nonhuman primates, sheep, dogs, pigs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal.

A "knock-in animal" is an animal (*e.g.*, a mammal such as a mouse or a rat) in which an endogenous gene has been substituted in part or in total with a heterologous gene (*i.e.*, a gene that is not endogenous to the locus in question; see Roamer *et al.*, *New Biol.* 1991, 3:331). This can be achieved by homologous recombination (see "knockout animal" below), transposition (Westphal and Leder, *Curr. Biol.* 1997; 7: 530), use of mutated recombination sites (Araki *et al.*, *Nucleic Acids Res.* 1997; 25: 868), PCR (Zhang and Henderson, *Biotechniques* 1998; 25: 784), or any other technique known in the art. The heterologous gene may be, *e.g.*, a reporter gene linked to the appropriate (*e.g.*, endogenous) promoter, which may be used to evaluate the expression or function of the endogenous gene (see, *e.g.*, Elegant *et al.*, *Proc. Natl. Acad. Sci. USA* 1998; 95: 11897).

A "knockout animal" is an animal (*e.g.*, a mammal such as a mouse or a rat) that has had a specific gene in its genome partially or completely inactivated by gene targeting (see, *e.g.*, U.S. Patents Nos. 5,777,195 and 5,616,491). A knockout animal can be a heterozygous knockout (*i.e.*, with one defective allele and one wild type allele) or a homozygous knockout (*i.e.*, with both alleles rendered defective). Preparation of a knockout animal typically requires first introducing a nucleic acid construct (a "knockout construct"), that will be used to decrease or eliminate expression of a particular gene, into an undifferentiated cell type termed an embryonic stem (ES) cell. The knockout construct is typically comprised of: (i) DNA from a portion (*e.g.*, an exon sequence, intron sequence, promoter sequence, or some combination thereof) of a gene to be knocked out; and (ii) a selectable marker sequence used to identify the presence of the knockout construct in the ES cell. The knockout construct is typically introduced (*e.g.*, electroporated) into ES cells so that it can homologously recombine with the genomic DNA of the cell in a double crossover

event. This recombined ES cell can be identified (*e.g.*, by Southern hybridization or PCR reactions that show the genomic alteration) and is then injected into a mammalian embryo at the blastocyst stage. In a preferred embodiment where the knockout animal is a mammal, a mammalian embryo with integrated ES cells is then implanted into a foster mother for the duration of gestation (see, *e.g.*, Zhou *et al.*, *Genes and Dev.* 1995; 9: 2623-34).

The phrases "disruption of the gene", "gene disruption", and the like, refer to: (i) insertion of a different or defective nucleic acid sequence into an endogenous (naturally occurring) DNA sequence, *e.g.*, into an exon or promoter region of a gene; or (ii) deletion of a portion of an endogenous DNA sequence of a gene; or (iii) a combination of insertion and deletion, so as to decrease or prevent the expression of that gene or its gene product in the cell as compared to the expression of the endogenous gene sequence.

In accordance with the present invention, there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. See, *e.g.*, Sambrook, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989 (herein "Sambrook *et al.*, 1989"); *DNA Cloning: A Practical Approach, Volumes I and II* (Glover ed. 1985); *Oligonucleotide Synthesis* (Gait ed. 1984); *Nucleic Acid Hybridization* (Hames and Higgins eds. 1985); *Transcription And Translation* (Hames and Higgins eds. 1984); *Animal Cell Culture* (Freshney ed. 1986); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); Ausubel *et al.* eds., *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. 1994; among others.

NPC1L1 Polynucleotides

The present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence encoding NPC1L1. More particularly, the present invention provides an isolated NPC1L1 nucleic acid sequence having a nucleotide sequence encoding mouse NPC1L1.

In one embodiment, the NPC1L1 nucleic acid has nucleotide sequence of SEQ ID NO:1, or a degenerate variant thereof. In another embodiment, NPC1L1 nucleic acid has nucleotide sequence of SEQ ID NO:2, or a degenerate variant thereof.

The present invention also provides an isolated single-stranded polynucleotide molecule comprising a nucleotide sequence that is the complement of a nucleotide sequence of one strand of any of the aforementioned nucleotide sequences (e.g., SEQ ID NO: 2).

The present invention further provides an isolated polynucleotide molecule comprising a nucleotide sequence that hybridizes to the complement of a polynucleotide that encodes the amino acid sequence of the mouse NPC1L1 protein of the present invention, under moderately stringent conditions, such as, for example, an aqueous solution of 2×SSC at 65°C; alternatively, for example, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% SDS, 1 mM EDTA at 65°C, and washing in 0.2 x SSC/0.1% SDS at 42°C (see the Definitions section above).

In a preferred embodiment, the homologous polynucleotide molecule hybridizes to the complement of a polynucleotide molecule comprising a nucleotide sequence that encodes the amino acid sequence of the mouse NPC1L1 protein of the present invention under highly stringent conditions, such as, for example, in an aqueous solution of 0.5×SSC at 65°C; alternatively, for example, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% SDS 1 mM EDTA at 65°C, and washing in 0.1.x SSC/0.1% SDS at 68°C (see the Definitions Section 5.1., above).

In a more preferred embodiment, the homologous polynucleotide molecule hybridizes under highly stringent conditions to the complement of a polynucleotide molecule consisting of a nucleotide sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:2.

The present invention further provides an isolated polynucleotide molecule comprising a nucleotide sequence that is homologous to the nucleotide sequence of a NPC1L1-encoding polynucleotide molecule of the present invention. In a preferred embodiment, such a polynucleotide molecule hybridizes under standard conditions to the complement of a polynucleotide molecule comprising a nucleotide sequence that

encodes the amino acid sequence of the mouse NPC1L1 protein of the present invention and has at least 75% sequence identity, preferably at least 80% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% sequence identity, and most preferably at least 99% sequence identity to the nucleotide sequence of such NPC1L1-encoding polynucleotide molecule (*e.g.*, as determined by a sequence comparison algorithm selected from BLAST, FASTA, DNA Strider, and GCG, and preferably as determined by the BLAST program from the National Center for Biotechnology Information (NCBI-Version 2.2), available on the WorldWideWeb at <www.ncbi.nlm.nih.gov/BLAST/htm>). In one embodiment, the homologous polynucleotide is homologous to a polynucleotide encoding mouse NPC1L1 protein of the present invention, e.g, SEQ ID NO: 2.

The present invention further provides an oligonucleotide molecule that hybridizes to a polynucleotide molecule of the present invention, or that hybridizes to a polynucleotide molecule having a nucleotide sequence that is the complement of a nucleotide sequence of a polynucleotide molecule of the present invention. Such an oligonucleotide molecule: (i) is about 10 nucleotides to about 200 nucleotides in length, preferably from about 15 to about 100 nucleotides in length, and more preferably about 20 to about 50 nucleotides in length, and (ii) hybridizes to one or more of the polynucleotide molecules of the present invention under highly stringent conditions (*e.g.*, washing in 6x SSC/0.5% sodium pyrophosphate at about 37°C for about 14-base oligos, at about 48°C for about 17-base oligos, at about 55°C for about 20-base oligos, and at about 60°C for about 23-base oligos). In one embodiment, an oligonucleotide molecule of the present invention is 100% complementary over its entire length to a portion of at least one of the aforementioned polynucleotide molecules of the present invention, and particularly any of SEQ ID NOs: 1 or 2. In another embodiment, an oligonucleotide molecule of the present invention is greater than 90% complementary over its entire length to a portion of at least one of the aforementioned polynucleotide molecules of the present invention, and particularly any of SEQ ID NOs: 1 or 2.

Specific non-limiting examples of oligonucleotide molecules according to the present invention include oligonucleotide molecules selected from the group consisting of SEQ ID NOs: 4 and 5.

Oligonucleotide molecules can be labeled, *e.g.*, with radioactive labels (*e.g.*, $\gamma^{32}P$), biotin, fluorescent labels, etc. In one embodiment, a labeled oligonucleotide molecule can be used as a probe to detect the presence of a nucleic acid. In another embodiment, two oligonucleotide molecules (one or both of which may be labeled) can be used as PCR primers, either for cloning a full-length nucleic acid or a fragment of a nucleic acid encoding a gene product of interest, or to detect the presence of nucleic acids encoding a gene product. Methods for conducting amplifications, such as the polymerase chain reaction (PCR), are described, among other places, in Saiki *et al.*, *Science* 1988, 239:487 and U.S. Patent No. 4,683,202. Other amplification techniques known in the art, *e.g.*, the ligase chain reaction, can alternatively be used (see, *e.g.*, U.S. Patent Nos. 6,335,184 and 6,027,923; Reyes *et al.*, *Clinical Chemistry* 2001; 47: 131-40; and Wu *et al.*, *Genomics* 1989; 4: 560-569).

The present invention further provides a polynucleotide molecule consisting of a nucleotide sequence that is a substantial portion of the nucleotide sequence of any of the aforementioned NPC1L1-related polynucleotide molecules of the present invention, or the complement of such nucleotide sequence. As used herein, a "substantial portion" of a NPC1L1-encoding nucleotide sequence means a nucleotide sequence that is less than the nucleotide sequence required to encode a complete NPC1L1 protein of the present invention, but comprising at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, or at least about 99% of the contiguous nucleotide sequence of a NPC1L1-encoding polynucleotide molecule of the present invention. Such polynucleotide molecules can be used for a variety of purposes including, *e.g.*, to express a portion of a NPC1L1 protein of the present invention in an appropriate expression system, or for use in conducting an assay to determine the expression level of a NPC1L1 gene in a biological sample, or to amplify a NPC1L1-encoding polynucleotide molecule.

In addition to the nucleotide sequences of any of the aforementioned NPC1L1-related polynucleotide molecules, polynucleotide molecules of the present invention can further comprise, or alternatively may consist of, nucleotide sequences selected from the sequence depicted in SEQ ID NO:1 (genomic) that naturally flank a NPC1L1-encoding nucleotide sequence in the chromosome, including regulatory sequences.

NPC1L1 Polypeptides

The present invention also provides an NPC1L1 polypeptide encoded by an NPC1L1 polynucleotide. In one embodiment, the NPC1L1 polypeptide is encoded by an NPC1L1 polynucleotide comprising the sequence as set forth in SEQ ID NO: 2.

The present invention also provides an NPC1L1 polypeptide encoded by an NPC1L1 polynucleotide that hybridizes to the complement of the polynucleotide sequence set forth in SEQ ID NOS. 1 or 2.

In one embodiment, NPC1L1 polypeptide comprises the amino acid sequence set forth SEQ ID NO:3.

The present invention further provides a non-human polypeptide that is homologous to the NPC1L1 protein of the present invention, as the term "homologous" is defined above for polypeptides. In one embodiment, the homologous NPC1L1 polypeptides of the present invention have the amino acid sequence identical to the amino acid sequence of SEQ ID NO:3, but have one or more amino acid residues conservatively substituted with a different amino acid residue. Conservative amino acid substitutions are well-known in the art. Rules for making such substitutions include those described by Dayhof, 1978, *Nat. Biomed. Res. Found.*, Washington, D.C., Vol. 5, Sup. 3, among others. More specifically, conservative amino acid substitutions are those that take place within a family of amino acids that are related in acidity, polarity, or bulkiness of their side chains. Genetically encoded amino acids are generally divided into four groups: (1) acidic=aspartate, glutamate; (2) basic=lysine, arginine, histidine; (3) non-polar=alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar=glycine, asparagine, glutamine, cysteine, serine,

threonine, tyrosine. Phenylalanine, tryptophan and tyrosine are also jointly classified as aromatic amino acids. One or more replacements within any particular group, *e.g.*, of a leucine with an isoleucine or valine, or of an aspartate with a glutamate, or of a threonine with a serine, or of any other amino acid residue with a structurally related amino acid residue, *e.g.*, an amino acid residue with similar acidity, polarity, bulkiness of side chain, or with similarity in some combination thereof, will generally have an insignificant effect on the function or immunogenicity of the polypeptide.

The NPC1L1 polypeptides of the present invention (including those encoded by the homologous polynucleotide molecules above, i.e., homologous NPC1L1 polypeptides) have the following functions including, but not limited to: (i) endocytosis and intracellular trafficking of multiple classes of lipids, including fatty acids such as oleic acid, sterols such as cholesterol, and, sphingolipids such as lactosylceramide; (ii) regulation of caveolae formation and/or internalization; (iii) the sensing of sterols through a sterol sensing domain; (iv) conferring localization to the ER and Golgi; and (v) regulating serum levels of total cholesterol, LDL-cholesterol, HDL-cholesterol, triglycerides, insulin, and glucose. (see also Davies et al., 2005, *J. Biological Chemistry,* Vol. 280, No. 13, pp. 12710-12720, the contents of which are expressly incorporated herein by reference).

Also encompassed by the present invention are orthologs of the specifically disclosed NPC1L1 polypeptides, and NPC1L1-encoding nucleic acids. Additional NPC1L1 orthologs can be identified based on the sequences of mouse and human orthologs disclosed herein, using standard sequence comparison algorithms such as BLAST, FASTA, DNA Strider, GCG, etc. In addition to mouse and human orthologs, particularly useful NPC1L1 orthologs of the present invention are monkey, dog, guinea pig, and porcine orthologs. As with the homologs discussed above, these orthologs can have the same functions as the NPC1L1 protein.

The present invention further provides a polypeptide consisting of a substantial portion of a mouse NPC1L1 protein of the present invention. "Substantial portion" has the same meaning as defined above under NPC1L1 polynucleotides.

The present invention further provides fusion proteins comprising any of the aforementioned polypeptides (proteins or peptide fragments) fused to a carrier or

fusion partner, as known in the art. For example, NPC1L1 can be fused with green fluorescent protein (GFP), V5, and Ig.

Recombinant Expression Systems Cloning and Expression Vectors

The present invention further provides compositions and constructs for cloning and expressing any of the NPC1L1 polynucleotide molecules of the present invention, including cloning vectors, expression vectors, transformed host cells comprising any of said vectors, and novel strains or cell lines derived therefrom. In one embodiment, the present invention provides a recombinant vector comprising a polynucleotide molecule having a nucleotide sequence encoding a non-human NPC1L1 polypeptide. In a specific embodiment, the mouse NPC1L1 polypeptide comprises the amino acid sequence of SEQ ID NO: 3.

Recombinant vectors of the present invention, particularly expression vectors, are preferably constructed so that the coding sequence for the NPC1L1 polynucleotide molecule of the present invention is in operative association with one or more regulatory elements necessary for transcription and translation of the coding sequence to produce a polypeptide. As used herein, the term "regulatory element" includes, but is not limited to, nucleotide sequences that encode inducible and non-inducible promoters, enhancers, operators and other elements known in the art that serve to drive and/or regulate expression of polynucleotide coding sequences. Also, as used herein, the coding sequence is in operative association with one or more regulatory elements where the regulatory elements effectively regulate and allow for the transcription of the coding sequence or the translation of its mRNA, or both.

Methods are known in the art for constructing recombinant vectors containing particular coding sequences in operative association with appropriate regulatory elements, and these can be used to practice the present invention. These methods include *in vitro* recombinant techniques, synthetic techniques, and *in vivo* genetic recombination. See, *e.g.*, the techniques described in Ausubel *et al.*, 1989, above; Sambrook *et al.*, 1989, above; Saiki *et al.*, 1988, above; Reyes *et al.*, 2001, above; Wu *et al.*, 1989, above; U.S. Patent Nos. 4,683,202; 6,335,184 and 6,027,923.

A variety of expression vectors are known in the art that can be utilized to express a polynucleotide molecule of the present invention, including recombinant bacteriophage DNA, plasmid DNA, and cosmid DNA expression vectors containing the particular coding sequences. Typical prokaryotic expression vector plasmids that can be engineered to contain a polynucleotide molecule of the present invention include pUC8, pUC9, pBR322 and pBR329 (Biorad Laboratories, Richmond, CA), pPL and pKK223 (Pharmacia, Piscataway, NJ), pQE50 (Qiagen, Chatsworth, CA), and pGEM-T EASY (Promega, Madison, WI), pcDNA6.2/V5-DEST and pcDNA3.2/V5DEST (Invitrogen, Carlsbad, CA) among many others. Typical eukaryotic expression vectors that can be engineered to contain a polynucleotide molecule of the present invention include an ecdysone-inducible mammalian expression system (Invitrogen, Carlsbad, CA), cytomegalovirus promoter-enhancer-based systems (Promega, Madison, WI; Stratagene, La Jolla, CA; Invitrogen), and baculovirus-based expression systems (Promega), among many others.

The regulatory elements of these and other vectors can vary in their strength and specificities. Depending on the host/vector system utilized, any of a number of suitable transcription and translation elements can be used. For instance, when cloning in mammalian cell systems, promoters isolated from the genome of mammalian cells, *e.g.*, mouse metallothionein promoter, or from viruses that grow in these cells, *e.g.*, vaccinia virus 7.5 K promoter or Maloney murine sarcoma virus long terminal repeat, can be used. Promoters obtained by recombinant DNA or synthetic techniques can also be used to provide for transcription of the inserted sequence. In addition, expression from certain promoters can be elevated in the presence of particular inducers, *e.g.*, zinc and cadmium ions for metallothionein promoters. Non-limiting examples of transcriptional regulatory regions or promoters include for bacteria, the β-gal promoter, the T7 promoter, the TAC promoter, λ left and right promoters, trp and lac promoters, trp-lac fusion promoters, etc.; for yeast, glycolytic enzyme promoters, such as ADH-I and -II promoters, GPK promoter, PGI promoter, TRP promoter, etc.; and for mammalian cells, SV40 early and late promoters, and adenovirus major late promoters, among others.

Specific initiation signals are also required for sufficient translation of inserted coding sequences. These signals typically include an ATG initiation codon and adjacent sequences. In cases where the polynucleotide molecule of the present invention, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translation control signals may be needed. However, in cases where only a portion of a coding sequence is inserted, exogenous translational control signals, including the ATG initiation codon, may be required. These exogenous translational control signals and initiation codons can be obtained from a variety of sources, both natural and synthetic. Furthermore, the initiation codon must be in-phase with the reading frame of the coding regions to ensure in-frame translation of the entire insert.

Expression vectors can also be constructed that will express a fusion protein comprising an NPC1L1 polypeptide of the present invention. Such fusion proteins can be used, *e.g.*, to raise anti-sera against a NPC1L1 polypeptide, to study the biochemical properties of the NPC1L1 polypeptide, to engineer a variant of a NPC1L1 polypeptide exhibiting different immunological or functional properties, or to aid in the identification or purification, or to improve the stability, of a recombinant NPC1L1 polypeptide. Possible fusion protein expression vectors include but are not limited to vectors incorporating sequences that encode β-galactosidase and trpE fusions, maltose-binding protein fusions, glutathione-S-transferase fusions, polyhistidine fusions (carrier regions), V5, HA, myc, and HIS. Methods known in the art can be used to construct expression vectors encoding these and other fusion proteins.

The fusion protein can be useful to aid in purification of the expressed protein. In non-limiting embodiments, *e.g.*, a NPC1L1-polyhistidine fusion protein can be purified using divalent nickel resin; a NPC1L1-maltose-binding fusion protein can be purified using amylose resin; and a NPC1L1-glutathione-S-transferase fusion protein can be purified using glutathione-agarose beads. Alternatively, antibodies against a carrier protein or peptide can be used for affinity chromatography purification of the fusion protein. For example, a nucleotide sequence coding for the target epitope of a monoclonal antibody can be engineered into the expression vector in operative

association with the regulatory elements and situated so that the expressed epitope is fused to a NPC1L1 protein of the present invention. In a non-limiting embodiment, a nucleotide sequence coding for the FLAG™ epitope tag (International Biotechnologies Inc.), which is a hydrophilic marker peptide, can be inserted by standard techniques into the expression vector at a point corresponding, *e.g.*, to the amino or carboxyl terminus of the NPC1L1 protein. The expressed NPC1L1 protein-FLAG™ epitope fusion product can then be detected and affinity-purified using commercially available anti-FLAG™ antibodies. The expression vector can also be engineered to contain polylinker sequences that encode specific protease cleavage sites so that the expressed NPC1L1 protein can be released from a carrier region or fusion partner by treatment with a specific protease. For example, the fusion protein vector can include a nucleotide sequence encoding a thrombin or factor Xa cleavage site, among others.

A signal sequence upstream from, and in reading frame with, the NPC1L1 coding sequence can be engineered into the expression vector by known methods to direct the trafficking and secretion of the expressed protein. Non-limiting examples of signal sequences include those from α-factor, immunoglobulins, outer membrane proteins, penicillinase, and T-cell receptors, among others.

To aid in the selection of host cells transformed or transfected with a recombinant vector of the present invention, the vector can be engineered to further comprise a coding sequence for a reporter gene product or other selectable marker. Such a coding sequence is preferably in operative association with the regulatory elements, as described above. Reporter genes that are useful in practicing the invention are known in the art, and include those encoding chloramphenicol acetyltransferase (CAT), green fluorescent protein and derivatives thereof, firefly luciferase, and human growth hormone, among others. Nucleotide sequences encoding selectable markers are known in the art, and include those that encode gene products conferring resistance to antibiotics or anti-metabolites, or that supply an auxotrophic requirement. Examples of such sequences include those that encode thymidine kinase activity, or resistance to methotrexate, ampicillin, kanamycin,

chloramphenicol, zeocin, pyrimethamine, aminoglycosides, hygromycin, blasticidine, or neomycin, among others.

Transformation of Host Cells

The present invention further provides a transformed host cell comprising a polynucleotide molecule or recombinant vector of the present invention, and a cell line derived therefrom. Such host cells are useful for cloning and/or expressing a polynucleotide molecule of the present invention. Such transformed host cells include but are not limited to microorganisms, such as bacteria transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA vectors, or yeast transformed with a recombinant vector, or animal cells, such as insect cells infected with a recombinant virus vector, *e.g.*, baculovirus, or mammalian cells infected with a recombinant virus vector, *e.g.*, adenovirus, vaccinia virus, lentivirus, adeno-associated virus (AAV), or herpesvirus, among others. For example, a strain of *E. coli* can be used such as, *e.g.*, the DH5α strain available from the ATCC, Manassas, VA, USA (Accession No. 31343), or from Stratagene (La Jolla, CA). Eukaryotic host cells include yeast cells, although mammalian cells, *e.g.*, from a mouse, rat, hamster, cow, monkey, or human cell line, among others, can also be utilized effectively. Examples of eukaryotic host cells that may be suitable for expressing a recombinant protein of the invention include Chinese hamster ovary (CHO) cells (*e.g.*, ATCC Accession No. CCL-61), NIH Swiss mouse embryo cells NIH/3T3 (*e.g.*, ATCC Accession No. CRL-1658), human epithelial kidney cells HEK 293 (*e.g.*, ATCC Accession No. CRL-1573), African green monkey COS-7 cells (ATCC Accession No. CRL-1651), human embryonal carcinoma NT2 cells (ATCC Accession No. CRL-1973), and human colon carcinoma Caco-2 cells ATCC Accession No. HTB-37.

The present invention provides for mammalian cells infected with a virus containing a recombinant viral vector of the present invention. For example, an overview and instructions concerning the infection of mammalian cells with adenovirus using the AdEasy™ Adenoviral Vector System is given in the Instructions Manual for this system from Stratagene (La Jolla, CA). As another example, an overview and instructions concerning the infection of mammalian cells with AAV

using the AAV Helper-Free System is given in the Instructions Manual for this system from Strategene (La Jolla, CA).

The recombinant vector of the invention is preferably transformed or transfected into one or more host cells of a substantially homogeneous culture of cells. The vector is generally introduced into host cells in accordance with known techniques, such as, *e.g.*, by protoplast transformation, calcium phosphate precipitation, calcium chloride treatment, microinjection, electroporation, transfection by contact with a recombined virus, liposome-mediated transfection, DEAE-dextran transfection, transduction, conjugation, or microprojectile bombardment, among others. Selection of transformants can be conducted by standard procedures, such as by selecting for cells expressing a selectable marker, *e.g.*, antibiotic resistance, associated with the recombinant expression vector.

Once an expression vector is introduced into the host cell, the presence of the polynucleotide molecule of the present invention, either integrated into the host cell genome or maintained episomally, can be confirmed by standard techniques, *e.g.*, by DNA-DNA, DNA-RNA, or RNA-antisense RNA hybridization analysis, restriction enzyme analysis, PCR analysis including reverse transcriptase PCR (RT-PCR), detecting the presence of a "marker" gene function, or by immunological or functional assay to detect the expected protein product.

Expression and Purification of Recombinant NPC1L1 Polypeptides

Once an NPC1L1 polynucleotide molecule of the present invention has been stably introduced into an appropriate host cell, the transformed host cell is clonally propagated, and the resulting cells can be grown under conditions conducive to the efficient production (*i.e.*, expression or overexpression) of the NPC1L1 polypeptide.

The polypeptide can be substantially purified or isolated from cell lysates, membrane fractions, or culture medium, as necessary, using standard methods, including but not limited to one or more of the following methods: ammonium sulfate precipitation, size fractionation, ion exchange chromatography, HPLC, density centrifugation, affinity chromatography, ethanol precipitation, and chromatofocusing.

During purification, the polypeptide can be detected based, *e.g.*, on size, or reactivity with a polypeptide-specific antibody, or by detecting the presence of a fusion tag.

For use in practicing the present invention, the polypeptide can be in an unpurified state as secreted into the culture fluid or as present in a cell lysate or membrane fraction. Alternatively, the polypeptide may be purified therefrom. Once a polypeptide of the present invention of sufficient purity has been obtained, it can be characterized by standard methods, including by SDS-PAGE, size exclusion chromatography, amino acid sequence analysis, immunological activity, biological activity, etc. The polypeptide can be further characterized using hydrophilicity analysis (see, *e.g.*, Hopp and Woods, *Proc. Natl. Acad. Sci. USA* 1981; 78: 3824), or analogous software algorithms, to identify hydrophobic and hydrophilic regions. Structural analysis can be carried out to identify regions of the polypeptide that assume specific secondary structures. Biophysical methods such as X-ray crystallography (Engstrom, *Biochem. Exp. Biol.* 1974; 11: 7-13), computer modeling (Fletterick and Zoller eds., *In: Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1986), and nuclear magnetic resonance (NMR) can be used to map and study potential sites of interaction between the polypeptide and other putative interacting proteins/receptors/molecules. Information obtained from these studies can be used to design deletion mutants, and to design or select therapeutic compounds that can specifically modulate the biological function of the NPC1L1 protein *in vivo*.

NPC1L1 Antibodies

The present invention also provides antibodies, including fragments thereof, which specifically bind to an NPC1L1 polypeptide, or fragment thereof. Antibodies to NPC1L1 have a number of applications, such as detecting the presence of NPC1L1 in a biological sample, determining the intracellular localization of NPC1L1, and modulating the activity of NPC1L1, *e.g.*, in a subject, for treatment (*e.g.*, therapeutic and prophylactic) of diseases and disorders associated with or mediated by NPC1L1, such as hyperlipidemia, obesity, type II diabetes, cardiovascular disease, and stroke. The present invention contemplates a number of sources for immunogenic NPC1L1 polypeptides for use in producing anti-NPC1L1 antibodies. These sources include

NPC1L1 polypeptides produced by recombinant technology and chemical synthesis; and products derived from their fragmentation or derivation.

Various antibodies against NPC1L1 are described in published U.S. patent application 2004/0161838, to Altmann et al., hereby incorporated by reference in its entirety. Such antibodies are designated A0715, A0716, A0717, A0718, A0867, A0868, A1801 or A1802. Additional commercially available antibodies include NPC1L1 rabbit polyclonal antibodies (Novus Biologicals, Littleton, CO, Cat # BC-400 NPC3).

As used herein, the term "antibody molecule" includes, but is not limited to, antibodies and binding fragments thereof, that specifically binds to an antigen, *e.g.*, an NPC1L1 protein. Suitable antibodies may be polyclonal (*e.g.*, sera or affinity purified preparations), monoclonal, or recombinant. Examples of useful fragments include separate heavy chains, light chains, Fab, $F(ab')_2$, Fabc, and Fv fragments. Fragments can be produced by enzymatic or chemical separation of intact immunoglobulins or by recombinant DNA techniques. Fragments may be expressed in the form of phage-coat fusion proteins (see, *e.g.*, International PCT Publication Nos. WO 91/17271, WO 92/01047, and WO 92/06204). Typically, the antibodies, fragments, or similar binding agents bind a specific antigen with an affinity of at least 10^7, 10^8, 10^9, or 10^{10} M^{-1}.

The present invention provides an isolated antibody directed against a polypeptide of the present invention. In a specific embodiment, antibodies can be raised against a NPC1L1 protein of the invention using known methods in view of this disclosure. Various host animals selected, *e.g.*, from pigs, cows, horses, rabbits, goats, sheep, rats, or mice, can be immunized with a partially or substantially purified NPC1L1 protein, or with a peptide homolog, fusion protein, peptide fragment, analog or derivative thereof, as described above. An adjuvant can be used to enhance antibody production.

Polyclonal antibodies can be obtained and isolated from the serum of an immunized animal and tested for specificity against the antigen using standard techniques. Alternatively, monoclonal antibodies can be prepared and isolated using

any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to; (i) the hybridoma technique originally described by Kohler and Milstein, *Nature* 1975; 256: 495-497; (ii) the trioma technique (Herring et al. (1988) Biomed. Biochim. Acta. 46:211-216 and Hagiwara et al. (1993) Hum. Antibod. Hybridomas 4:15); (iii) the human B-cell hybridoma technique (Kosbor *et al.*, *Immunology Today* 1983; 4: 72; Cote *et al.*, *Proc. Natl. Acad. Sci. USA* 1983; 80: 2026-2030); and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., 1985, pp. 77-96). Alternatively, techniques described for the production of single chain antibodies (see, *e.g.*, U.S. Patent No. 4,946,778) can be adapted to produce NPC1L1-specific single chain antibodies.

Antibody fragments that contain specific binding sites for the NPC1L1 polypeptide of the present invention are also encompassed within the present invention, and can be generated by known techniques. Such fragments include but are not limited to $F(ab')_2$ fragments, which can be generated by pepsin digestion of an intact antibody molecule, and Fab fragments, which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries can be constructed (Huse *et al.*, *Science* 1989; 246: 1275-1281) to allow rapid identification of Fab fragments having the desired specificity to the particular NPC1L1 protein.

Techniques for the production and isolation of monoclonal antibodies and antibody fragments are known in the art, and are generally described, among other places, in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, and in Goding, *Monoclonal Antibodies: Principles and Practice*, Academic Press, London, 1986. The art also provides recombinant expression systems in bacteria and yeast, enabling the production of functional antibodies that are analogous to those normally found in vertebrate systems. (Skerra et al. (1988) Science 240:1038-1041, Better et al. (1988) Science 240:1041-1043, and Bird et al. (1988) Science 242:423-426, Horwitz et al. (1989) Proc. Natl. Acad. Sci. USA. 85:8678-82.)

Antibodies or antibody fragments can be used in methods known in the art relating to the localization and activity of NPC1L1, *e.g.*, in Western blotting, *in situ*

imaging, measuring levels thereof in appropriate physiological samples, etc. Immunoassay techniques using antibodies include radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, *in situ* immunoassays (using, *e.g.*, colloidal gold, enzyme or radioisotope labels), precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. Antibodies can also be used in microarrays (see, *e.g.*, International PCT Publication No. WO 00/04389). Furthermore, antibodies can be used as therapeutics to inhibit the activity of a NPC1L1 protein.

Recent advances in antibody engineering have allowed the genes encoding antibodies to be manipulated, so that antigen-binding molecules can be expressed within mammalian cells. Application of gene technologies to antibody engineering has enabled the synthesis of single-chain fragment variable (scFv) antibodies that combine within a single polypeptide chain the light and heavy chain variable domains of an antibody molecule covalently joined by a pre-designed peptide linker. Intracellular antibody (or "intrabody") strategy serves to target molecules involved in essential cellular pathways for modification or ablation of protein function. Antibody genes for intracellular expression can be derived, *e.g.*, either from murine or human monoclonal antibodies or from phage display libraries. For intracellular expression, small recombinant antibody fragments containing the antigen recognizing and binding regions can be used. Intrabodies can be directed to different intracellular compartments by targeting sequences attached to the antibody fragments.

Various methods have been developed to produce intrabodies. Techniques described for the production of single chain antibodies (see, *e.g.*, U.S. Patents No. 5,476,786; 5,132,405; and 4,946,778) can be adapted to produce polypeptide-specific single chain antibodies. Another method called intracellular antibody capture (IAC), is based on a genetic screening approach (Tanaka *et al.*, *Nucleic Acids Res.* 2003; 31: e23). Using this technique, consensus immunoglobulin variable frameworks are identified that can form the basis of intrabody libraries for direct screening.. The

procedure comprises *in vitro* production of a single antibody gene fragment from oligonucleotides and diversification of CDRs of the immunoglobulin variable domain by mutagenic PCR to generate intrabody libraries. This method obviates the need for *in vitro* production of antigen for pre-selection of antibody fragments, and also yields intrabodies with enhanced intracellular stability.

Intrabodies can be used to modulate cellular physiology and metabolism through a variety of mechanisms, including blocking, stabilizing, or mimicking protein-protein interactions, by altering enzyme function, or by diverting proteins from their usual intracellular compartments. Intrabodies can be directed to the relevant cellular compartments by modifying the genes that encode them to specify N- or C-terminal polypeptide extensions for providing intracellular-trafficking signals.

NPC1L1 Applications

NPC1L1 polynucleotides and polypeptides of the present invention are useful for a variety of purposes, including for use in cell-based or non-cell-based assays to identify molecules that interact with NPC1L1 relevant to its *in vivo* function, to screen for compounds that bind to NPC1L1 and modulate its expression and/or activity and are therefore useful as therapeutic compounds to treat or prevent NPC1L1-mediated diseases or disorders as described herein, or as antigens to raise polyclonal or monoclonal antibodies, as described below. Such antibodies can be used as therapeutic agents to modulate the activity of NPC1L1 activity, or as diagnostic reagents, *e.g.*, using standard techniques such as Western blot assays or immunostaining, to screen for NPC1L1 protein expression levels in cell, tissue or fluid samples collected from a subject.

A polypeptide of the present invention can be modified at the protein level to improve or otherwise alter its biological or immunological characteristics. One or more chemical modifications of the polypeptide can be carried out using known techniques to prepare analogs therefrom, including but not limited to any of the following: substitution of one or more L-amino acids of the polypeptide with corresponding D-amino acids, amino acid analogs, or amino acid mimics, so as to

produce, *e.g.*, carbazates or tertiary centers; or specific chemical modification, such as, *e.g.*, proteolytic cleavage with trypsin, chymotrypsin, papain or V8 protease, or treatment with $NaBH_4$ or cyanogen bromide, or acetylation, formylation, oxidation or reduction, etc. Alternatively or additionally, a polypeptide of the present invention can be modified by genetic recombination techniques.

A polypeptide of the present invention can be derivatized, by conjugation thereto of one or more chemical groups, including but not limited to acetyl groups, sulfur bridging groups, glycosyl groups, lipids, and phosphates, and/or by conjugation to a second polypeptide of the present invention, or to another protein, such as, *e.g.*, serum albumin, keyhole limpet hemocyanin, or commercially activated BSA, or to a polyamino acid (*e.g.*, polylysine), or to a polysaccharide, (*e.g.*, sepharose, agarose, or modified or unmodified celluloses), among others. Such conjugation is preferably by covalent linkage at amino acid side chains and/or at the N-terminus or C-terminus of the polypeptide. Methods for carrying out such conjugation reactions are known in the field of protein chemistry.

Derivatives useful in practicing the claimed invention also include those in which a water-soluble polymer such as, *e.g.*, polyethylene glycol, is conjugated to a polypeptide of the present invention, or to an analog or derivative thereof, thereby providing additional desirable properties while retaining, at least in part, the immunogenicity of the polypeptide. These additional desirable properties include, *e.g.*, increased solubility in aqueous solutions, increased stability in storage, increased resistance to proteolytic degradation, and increased *in vivo* half-life. Water-soluble polymers suitable for conjugation to a polypeptide of the present invention include but are not limited to polyethylene glycol homopolymers, polypropylene glycol homopolymers, copolymers of ethylene glycol with propylene glycol, wherein said homopolymers and copolymers are unsubstituted or substituted at one end with an alkyl group, polyoxyethylated polyols, polyvinyl alcohol, polysaccharides, polyvinyl ethyl ethers, and α,β-poly[2-hydroxyethyl]-DL-aspartamide. Polyethylene glycol is particularly preferred. Methods for making water-soluble polymer conjugates of polypeptides are known in the art and are described, among other places, in U.S. Patent Nos. 3,788,948; 3,960,830; 4,002,531; 4,055,635; 4,179,337; 4,261,973;

4,412,989; 4,414,147; 4,415,665; 4,609,546; 4,732,863; and 4,745,180; European Patent (EP) 152,847; EP 98,110; and Japanese Patent 5,792,435; which patents are incorporated herein by reference.

Targeted Mutation of the NPC1L1 Gene

Based on the present disclosure of polynucleotide molecules, genetic constructs can be prepared for use in disabling or otherwise mutating a mammalian NPC1L1 gene. For example, the mouse NPC1L1 gene can be mutated using an appropriately designed genetic construct in combination with genetic techniques currently known or to be developed in the future. In another instance, the mouse NPC1L1 gene can be mutated using a genetic construct that functions to: (i) delete all or a portion of the coding sequence or regulatory sequence of the NPC1L1 gene; (ii) replace all or a portion of the coding sequence or regulatory sequence of the NPC1L1 gene with a different nucleotide sequence; (iii) insert into the coding sequence or regulatory sequence of the NPC1L1 gene one or more nucleotides, or an oligonucleotide molecule, or polynucleotide molecule, which can comprise a nucleotide sequence from the same species or from a heterologous source; or (iv) carry out some combination of (i), (ii) and (iii).

Cells, tissues and animals that are mutated for the NPC1L1 gene are useful for a number of purposes, such as further studying the biological function of NPC1L1, and conducting screens to identify therapeutic compounds that selectively modulate NPC1L1 expression and/or activity. In a preferred embodiment, the mutation serves to partially or completely disable the NPC1L1 gene, or partially or completely disable the protein encoded by the NPC1L1 gene. In this context, a NPC1L1 gene or protein is considered to be partially or completely disabled if either no protein product is made (for example, where the gene is deleted), or a protein product is made that can no longer carry out its normal biological function or can no longer be transported to its normal cellular location, or a protein product is made that carries out its normal biological function but at a significantly reduced level.

In a non-limiting embodiment, a genetic construct of the present invention is used to mutate a wild-type NPC1L1 gene by replacement of at least a portion of the

coding or regulatory sequence of the wild-type gene with a different nucleotide sequence such as, *e.g.*, a mutated coding sequence or mutated regulatory region, or portion thereof. A mutated NPC1L1 gene sequence for use in such a genetic construct can be produced by any of a variety of known methods, including by use of error-prone PCR, or by cassette mutagenesis. For example, oligonucleotide-directed mutagenesis can be employed to alter the coding or regulatory sequence of a wild-type NPC1L1 gene in a defined way, *e.g.*, to introduce a frame-shift or a termination codon at a specific point within the sequence. A mutated nucleotide sequence for use in the genetic construct of the present invention can be prepared by insertion into the coding or regulatory (*e.g.*, promoter) sequence of one or more nucleotides, oligonucleotide molecules or polynucleotide molecules, or by replacement of a portion of the coding sequence or regulatory sequence with one or more different nucleotides, oligonucleotide molecules or polynucleotide molecules. Such oligonucleotide molecules or polynucleotide molecules can be obtained from any naturally occurring source or can be synthetic. The inserted sequence can serve simply to disrupt the reading frame of the NPC1L1 gene, or can further encode a heterologous gene product such as a selectable marker.

In one embodiment, NPC1LI can be mutated in the transmembrane-spanning region, putative sterol sensing domain, amino-terminal 'NPC1 domain' domain, and/or ER/Goli targeting signal.

Mutations to produce modified cells, tissues and animals that are useful in practicing the present invention can occur anywhere in the NPC1L1 gene, including the open reading frame, the promoter or other regulatory region, or any other portion of the sequence that naturally comprises the gene or ORF. Such cells include mutants in which a modified form of the NPC1L1 protein normally encoded by the NPC1L1 gene is produced, or in which no protein normally encoded by the NPC1L1 gene is produced. Such cells can be null, conditional or leaky mutants.

Alternatively, a genetic construct can comprise nucleotide sequences that naturally flank the NPC1L1 gene or ORF *in situ*, with only a portion or no nucleotide

sequences from the actual coding region of the gene itself. Such a genetic construct can be useful to delete the entire NPC1L1 gene or ORF.

Methods for carrying out homologous gene replacement are known in the art. For targeted gene mutation through homologous recombination, the genetic construct is preferably a plasmid, either circular or linearized, comprising a mutated nucleotide sequence as described above. In a non-limiting embodiment, at least about 200 nucleotides of the mutated sequence are used to specifically direct the genetic construct of the present invention to the particular targeted NPC1L1 gene for homologous recombination, although shorter lengths of nucleotides may also be effective. In addition, the plasmid preferably comprises an additional nucleotide sequence encoding a reporter gene product or other selectable marker constructed so that it will insert into the genome in operative association with the regulatory element sequences of the native NPC1L1 gene to be disrupted. Reporter genes that can be used in practicing the invention are known in the art, and include those encoding CAT, green fluorescent protein, and β-galactosidase, among others. Nucleotide sequences encoding selectable markers are also known in the art, and include those that encode gene products conferring resistance to antibiotics or anti-metabolites, or that supply an auxotrophic requirement.

In view of the present disclosure, methods that can be used for creating the genetic constructs of the present invention will be apparent, and can include *in vitro* recombinant techniques, synthetic techniques, and *in vivo* genetic recombination, as described, among other places, in Ausubel *et al.*, 1989, above; Sambrook *et al.*, 1989, above; Innis *et al.*, 1995, above; and Erlich, 1992, above.

Mammalian cells can be transformed with a genetic construct of the present invention in accordance with known techniques, such as, *e.g.*, by electroporation. Selection of transformants can be carried out using standard techniques, such as by selecting for cells expressing a selectable marker associated with the construct. Identification of transformants in which a successful recombination event has occurred and the particular target gene has been disabled can be carried out by genetic analysis, such as by Southern blot analysis, or by Northern analysis to detect a lack of mRNA transcripts encoding the particular protein, or by the appearance of cells

lacking the particular protein, as determined, *e.g.*, by immunological analysis, or some combination thereof.

The present invention thus provides modified mammalian cells in which the native NPC1L1 gene has been mutated. The present invention further provides modified animals in which the NPC1L1 gene has been mutated.

Genetically Modified Animals

Genetically modified animals can be produced for studying the biological function of the NPC1L1 of the present invention *in vivo* and for screening and/or testing candidate compounds, *e.g.*, inhibitors, such as antisense nucleic acids, shRNAs, siRNAs, or ribozymes, small molecules, or antibodies, for their ability to affect, *e.g.*, inhibit, the expression and/or activity of NPC1L1 as potential therapeutics for treating disorders of lipid metabolism, such as hyperlipidemia, *e.g.*, hypercholesterolemia, obesity, type II diabetes, cardiovascular disease, and stroke. Other candidate compounds, *e.g.*, NPC1L1 agonists, may be identified and/or tested for their ability to enhance or increase the expression and/or activity of NPC1L1 as potential therapeutics for treating disorders such as anorexia, cachexia, and wasting, using the genetically modified animals described herein.

To investigate the function of NPC1L1 *in vivo* in animals, NPC1L1-encoding polynucleotides or NPC1L1-inhibiting antisense nucleic acids, shRNAs, siRNAs, or ribozymes can be introduced into test animals, such as mice or rats, using, *e.g.*, viral vectors or naked nucleic acids. Alternatively, transgenic animals can be produced. Specifically, "knock-in" animals with the endogenous NPC1L1 gene substituted with a heterologous gene or an ortholog from another species or a mutated NPC1L1 gene, or "knockout" animals with NPC1L1 gene partially or completely inactivated, or transgenic animals expressing or overexpressing a wild-type or mutated NPC1L1 gene (*e.g.*, upon targeted or random integration into the genome) can be generated.

NPC1L1-encoding nucleic acids can be introduced into animals using viral delivery systems. Exemplary viruses for production of delivery vectors include

without limitation adenovirus, herpesvirus, retroviruses, vaccinia virus, and adeno-associated virus (AAV). See, *e.g.*, Becker *et al.*, *Meth. Cell Biol.* 1994; 43: 161-89; Douglas and Curiel, *Science & Medicine* 1997; 4: 44-53; Yeh and Perricaudet, *FASEB J.* 1997; 11: 615-623; Kuo *et al.*, *Blood* 1993; 82: 845; Markowitz *et al.*, *J. Virol.* 1988; 62: 1120; Mann *et al.*, *Cell* 1983; 33: 153; U.S. Patents No. 5,399,346; 4,650,764; 4,980,289; 5,124,263; and International Publication No. WO 95/07358.

In an alternative method, a NPC1L1-encoding nucleic acid can be introduced by liposome-mediated transfection, a technique that provides certain practical advantages, including the molecular targeting of liposomes to specific cells. Directing transfection to particular cell types (also possible with viral vectors) is particularly advantageous in a tissue with cellular heterogeneity, such as the brain, pancreas, liver, and kidney. Lipids may be chemically coupled to other molecules for the purpose of targeting. Targeted peptides (*e.g.*, hormones or neurotransmitters), proteins such as antibodies, or non-peptide molecules can be coupled to liposomes chemically.

In another embodiment, target cells can be removed from an animal, and a nucleic acid can be introduced as a naked construct. The transformed cells can be then re-implanted into the body of the animal. Naked nucleic acid constructs can be introduced into the desired host cells by methods known in the art, *e.g.*, transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, use of a gene gun or use of a DNA vector transporter. See, *e.g.*, Wu et al., J. Biol. Chem. 1992; 267: 963-7; Wu et al., J. Biol. Chem. 1988; 263: 14621-4.

In yet another embodiment, NPC1L1-encoding nucleic acids can be introduced into animals by injecting naked plasmid DNA containing a NPC1L1-encoding nucleic acid sequence into the tail vein of animals, in particular mammals (Zhang et al., Hum. Gen. Ther. 1999, 10:1735-7). This injection technique can also be used to introduce siRNA targeted to NPC1L1 into animals, in particular mammals (Lewis *et al.*, *Nature Genetics* 2002, 32: 105-106).

As specified above, transgenic animals can also be generated. Methods of making transgenic animals are well-known in the art (for transgenic mice see *Gene Targeting: A Practical Approach*, 2nd Ed., Joyner ed., IRL Press at Oxford University Press, New York, 2000; *Manipulating the Mouse Embryo: A Laboratory Manual*, Nagy *et al.* eds., Cold Spring Harbor Press, New York, 2003; *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson ed., IRL Press at Oxford University Press, 1987; *Transgenic Animal Technology: A Laboratory Handbook*, Pinkert ed., Academic Press, New York, 1994; Hogan, *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1986; Brinster *et al.*, *Proc. Nat. Acad. Sci. USA* 1985; 82: 4438- 4442; Capecchi, *Science* 1989; 244: 1288-1292; Joyner *et al.*, *Nature* 1989; 338: 153-156; U.S. Patents No. 4,736,866; 4,870,009; 4,873,191; for particle bombardment see U.S. Patent No. 4, 945,050; for transgenic rats see, *e.g.*, Hammer *et al.*, *Cell* 1990; 63: 1099-1112; for non-rodent transgenic mammals and other animals see, *e.g.*, Pursel *et al.*, *Science* 1989; 244: 1281-1288 and Simms *et al.*, *Bio/Technology* 1988; 6: 179- 183; and for culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation and direct injection see, *e.g.*, *Teratocarcinomas and Embryonic Stem Cells, A Practical Approach*, Robertson ed., IRL Press, 1987). Clones of the nonhuman transgenic animals can be produced according to available methods (see *e.g.*, Wilmut *et al.*, *Nature* 1997; 385: 810-813 and International Publications No. WO 97/07668 and WO 97/07669).

In one embodiment, the transgenic animal is a "knockout" animal having a heterozygous or homozygous alteration in the sequence of an endogenous NPC1L1 gene that results in a decrease of NPC1L1 function, preferably such that NPC1L1 expression is undetectable or insignificant. Knockout animals are typically generated by homologous recombination with a vector comprising a transgene having at least a portion of the gene to be knocked out. Typically a deletion, addition or substitution has been introduced into the transgene to functionally disrupt it.

Knockout animals can be prepared by any method known in the art (see, *e.g.*, Snouwaert *et al.*, *Science* 1992; 257: 1083; Lowell *et al.*, *Nature* 1993; 366: 740-42;

Capecchi, *Science* 1989; 244: 1288-1292; Palmiter *et al.*, *Ann. Rev. Genet.* 1986; 20: 465-499; Bradley, . *Current Opinion . in Bio/Technology* 1991; 2: 823-829; and International Publications No. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169). Preparation of a knockout animal typically requires first introducing a nucleic acid construct (a "knockout construct"), that will be used to decrease or eliminate expression of a particular gene, into an undifferentiated cell type termed an embryonic stem (ES) cell. The knockout construct is typically comprised of: (i) DNA from a portion (*e.g.*, an exon sequence, intron sequence, promoter sequence, or some combination thereof) of a gene to be knocked out; and (ii) a selectable marker sequence used to identify the presence of the knockout construct in the ES cell. The knockout construct is typically introduced (*e.g.*, electroporated or microinjected) into ES cells so that it can homologously recombine with the genomic DNA of the cell in a double crossover event. This recombined ES cell can be identified (*e.g.*, by Southern hybridization or PCR reactions that show the genomic alteration) and is then injected into a mammalian embryo at the blastocyst stage. In a preferred embodiment where the knockout animal is a mammal, a mammalian embryo with integrated ES cells is then implanted into a foster mother for the duration of gestation (see, *e.g.*, Zhou *et al.*, *Genes and Dev.* 1995; 9: 2623-34).

In a specific embodiment, the knockout vector is designed such that, upon homologous recombination, the endogenous NPC1L1-related gene is functionally disrupted (*i.e.*, no longer encodes a functional protein). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NPC1L1-related gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the NPC1L1-related polypeptide). In the homologous recombination vector, the altered portion of NPC1L1-related gene is preferably flanked at its 5' and 3' ends by additional nucleic acid of the NPC1L1-related gene to allow for homologous recombination to occur between the exogenous NPC1L1-related gene carried by the vector and an endogenous NPC1L1-related gene in an embryonic stem cell. The additional flanking NPC1L1-related nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (at both the 5' and 3' ends) are included in the vector (see, *e.g.*, Thomas

and Capecchi, *Cell* 1987; 51: 503). The vector is introduced into an ES cell line (*e.g.*, by electroporation), and cells in which the introduced NPC1L1-related gene has homologously recombined with the endogenous NPC1L1-related gene are selected (see, *e.g.*, Li *et al.*, *Cell* 1992; 69: 915). The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see, *e g.*, Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson ed., IRL, Oxford, 1987, pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene.

The phenotype of knockout animals can be predictive of the *in vivo* function of the gene and of the effects or lack of effect of its antagonists or agonists. Knockout animals can also be used to study the effects of the NPC1L1 protein in models of disease, including, hyperlipidemia and other lipid-mediated disorders. In a specific embodiment, knockout animals, such as mice harboring the NPC1L1 gene knockout, may be used to produce antibodies against the heterologous NPC1L1 protein (*e.g.*, human NPC1L1) (Claesson *et al.*, *Scan. J. Immunol.* 1994; 0: 257-264; Declerck *et al.*, *J. Biol. Chem.* 1995; 270: 8397-400).

Genetically modified animals expressing or harboring NPC1L1-specific antisense polynucleotides, shRNA, siRNA, or ribozymes can be used analogously to knockout animals described above.

In another embodiment of the invention, the transgenic animal is an animal having an alteration in its genome that results in altered expression (*e.g.*, increased or decreased expression) of the NPC1L1 gene, *e.g.*, by introduction of additional copies of NPC1L1 gene in various parts of the genome, or by operatively inserting a regulatory sequence that provides for altered expression of an endogenous copy of the NPC1L1 gene. Such regulatory sequences include inducible, tissue-specific, and constitutive promoters and enhancer elements. Suitable promoters include metallothionein, albumin (Pinkert *et al.*, *Genes Dev.* 1987; 1: 268-76), and K-14 keratinocyte (Vassar *et al.*, *Proc. Natl. Acad. Sci. USA* 1989; 86: 1563-1567) gene

promoters. Overexpression or underexpression of the wild-type NPC1L1 polypeptide, polypeptide fragment or a mutated version thereof may alter normal cellular processes, resulting in a phenotype that identifies a tissue in which NPC1L1 expression is functionally relevant and may indicate a therapeutic target for the NPC1L1, its agonists or antagonists. For example, a transgenic test animal can be engineered to overexpress or underexpress a full-length NPC1L1 sequence, which may result in a phenotype that shows similarity with human diseases.

Transgenic animals can also be produced that allow for regulated (*e.g.*, tissue-specific) expression of the transgene. One example of such a system that may be produced is the Cre-Lox recombinase system of bacteriophage P1 (Lakso *et al.*, *Proc. Natl. Acad. Sci. USA* 1992; 89: 6232-6236; U.S. Patents No. 4,959,317 and 5,801,030). If the Cre-Lox recombinase system is used to regulate expression of a transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic or gene-targeted animals, one containing a transgene encoding a selected protein or containing a targeted allele (*e.g.*, a loxP flanked exon), and the other containing a transgene encoding a recombinase (*e.g.*, a tissue-specific expression of Cre recombinase). Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.*, *Science* 1991; 251: 1351-1355; U.S. Patent No. 5,654,182). In another embodiment, both Cre-Lox and Flp-Frt are used in the same system to regulate expression of the transgene, and for sequential deletion of vector sequences in the same cell (Sun *et al.*, Nat. Genet. 2000; 25: 83-6). Regulated transgenic animals can be also prepared using the tet-repressor system (see, *e.g.*, U.S. Patent No. 5,654,168).

The *in vivo* function of NPC1L1 can be also investigated through making "knock-in" animals. In such animals the endogenous NPC1L1 gene can be replaced, *e.g.*, by a heterologous gene, by a NPC1L1 ortholog or by a mutated NPC1L1 gene. See, for example, Wang *et al.*, *Development* 1997; 124: 2507-2513; Zhuang *et al.*, *Mol. Cell Biol.* 1998; 18: 3340-3349; Geng *et al.*, *Cell* 1999; 97: 767-777; Baudoin *et al.*, *Genes Dev.* 1998; 12: 1202-1216. Thus, a non-human transgenic animal can be

created in which: (i) a human ortholog of the non-human animal NPC1L1 gene has been stably inserted into the genome of the animal; and/or (ii) the endogenous non-human animal NPC1L1 gene has been replaced with its human counterpart (see, *e.g.*, Coffman, *Semin. Nephrol.* 1997; 17: 404; Esther *et al.*, *Lab. Invest.* 1996; 74: 953; Murakami *et al.*, *Blood Press. Suppl.* 1996; 2: 36). In one aspect of this embodiment, a human NPC1L1 gene inserted into the transgenic animal is the wild-type human NPC1L1 gene. In another aspect, the NPC1L1 gene inserted into the transgenic animal is a mutated form or a variant of the human NPC1L1 gene.

Included within the scope of the present invention are transgenic animals, preferably mammals (*e.g.*, mice) in which, in addition to the NPC1L1 gene, one or more additional genes (preferably, associated with hyperlipidemia or related disorders) have been knocked out, or knocked in, or overexpressed. Such animals can be generated by repeating the procedures set forth herein for generating each construct, or by breeding two animals of the same species (each with a different single gene manipulated) to each other, and screening for those progeny animals having the desired genotype.

Inhibition of NPC1L1

As specified above, the NPC1L1-encoding nucleic acid molecules of the can be used to inhibit the expression of NPC1L1 genes (*e.g.*, by inhibiting transcription, splicing, transport, or translation or by promoting degradation of corresponding mRNAs). Specifically, the nucleic acid molecules of the invention can be used to "knock down" or "knock out" the expression of the NPC1L1 genes in a cell or tissue (*e.g.*, in an animal model or in cultured cells) by using their sequences to design antisense oligonucleotides, RNA interference (RNAi) molecules, ribozymes, nucleic acid molecules to be used in triplex helix formation, etc. Preferred methods to inhibit gene expression are described below.

In one embodiment the transcription of NPC1L1 mRNA is inhibited by targeting NPC1L1 promoter transcription factors using an agonist or antagonist to these factors. In this embodiment the specific agonist or antagonist is identified by its ability to downregulate the expression of a reporter gene (such as luciferase or green

fluorescence protein) driven by the promoter for NPC1L1, *e.g.*, the mouse, rat or human promoter.

RNA Interference (RNAi). RNA interference (RNAi) is a process of sequence-specific post-transcriptional gene silencing by which double stranded RNA (dsRNA) homologous to a target locus can specifically inactivate gene function in plants, fungi, invertebrates, and vertebrates, including mammals (Hammond *et al.*, *Nature Genet.* 2001; 2: 110-119; Sharp, *Genes Dev.* 1999;13: 139-141). This dsRNA-induced gene silencing is mediated by short double-stranded small interfering RNAs (siRNAs) generated from longer dsRNAs by ribonuclease III cleavage (Bernstein *et al.*, *Nature* 2001; 409: 363-366 and Elbashir *et al.*, *Genes Dev.* 2001; 15: 188-200). RNAi-mediated gene silencing is thought to occur via sequence-specific mRNA degradation, where sequence specificity is determined by the interaction of an siRNA with its complementary sequence within a target mRNA (see, *e.g.*, Tuschl, *Chem. Biochem.* 2001; 2: 239-245).

For mammalian systems, RNAi commonly involves the use of dsRNAs that are greater than 500 bp; however, it can also be activated by introduction of either siRNAs (Elbashir, *et al.*, *Nature* 2001; 411: 494-498) or short hairpin RNAs (shRNAs) bearing a fold back stem-loop structure (Paddison *et al.*, *Genes Dev.* 2002; 16: 948-958; Sui *et al.*, *Proc. Natl. Acad. Sci. USA* 2002; 99: 5515-5520; Brummelkamp *et al.*, *Science* 2002; 296: 550-553; Paul *et al.*, *Nature Biotechnol.* 2002; 20: 505-508).

The siRNAs to be used in the methods of the present invention are preferably short double stranded nucleic acid duplexes comprising annealed complementary single stranded nucleic acid molecules. In preferred embodiments, the siRNAs are short dsRNAs comprising annealed complementary single strand RNAs. However, the invention also encompasses embodiments in which the siRNAs comprise an annealed RNA:DNA duplex, wherein the sense strand of the duplex is a DNA molecule and the antisense strand of the duplex is a RNA molecule. In one embodiment, an siRNA of the invention is set forth as SEQ ID NO: 23 or SEQ ID NO: 24.

Preferably, each single stranded nucleic acid molecule of the siRNA duplex is of from about 19 nucleotides to about 27 nucleotides in length. In preferred embodiments, duplexed siRNAs have a 2 or 3 nucleotide 3' overhang on each strand of the duplex. In preferred embodiments, siRNAs have 5'-phosphate and 3'-hydroxyl groups.

The RNAi molecules to be used in the methods of the present invention comprise nucleic acid sequences that are complementary to the nucleic acid sequence of a portion of the target locus. In certain embodiments, the portion of the target locus to which the RNAi probe is complementary is at least about 15 nucleotides in length. In preferred embodiments, the portion of the target locus to which the RNAi probe is complementary is at least about 19 nucleotides in length. The target locus to which an RNAi probe is complementary may represent a transcribed portion of the NPC1L1 gene or an untranscribed portion of the NPC1L1 gene (*e.g.*, intergenic regions, repeat elements, etc.).

The RNAi molecules may include one or more modifications, either to the phosphate-sugar backbone or to the nucleoside. For example, the phosphodiester linkages of natural RNA may be modified to include at least one heteroatom other than oxygen, such as nitrogen or sulfur. In this case, for example, the phosphodiester linkage may be replaced by a phosphothioester linkage. Similarly, bases may be modified to block the activity of adenosine deaminase. Where the RNAi molecule is produced synthetically, or by *in vitro* transcription, a modified ribonucleoside may be introduced during synthesis or transcription.

According to the present invention, siRNAs may be introduced to a target cell as an annealed duplex siRNA, or as single stranded sense and anti-sense nucleic acid sequences that, once within the target cell, anneal to form the siRNA duplex. Alternatively, the sense and anti-sense strands of the siRNA may be encoded on an expression construct that is introduced to the target cell. Upon expression within the target cell, the transcribed sense and antisense strands may anneal to reconstitute the siRNA.

The shRNAs to be used in the methods of the present invention comprise a single stranded "loop" region connecting complementary inverted repeat sequences that anneal to form a double stranded "stem" region. Structural considerations for shRNA design are discussed, for example, in McManus *et al.*, *RNA* 2002; 8: 842-850. In certain embodiments the shRNA may be a portion of a larger RNA molecule, *e.g.*, as part of a larger RNA that also contains U6 RNA sequences (Paul *et al.*, *supra*).

In preferred embodiments, the loop of the shRNA is from about 1 to about 9 nucleotides in length. In preferred embodiments the double stranded stem of the shRNA is from about 19 to about 33 base pairs in length. In preferred embodiments, the 3' end of the shRNA stem has a 3' overhang. In particularly preferred embodiments, the 3' overhang of the shRNA stem is from 1 to about 4 nucleotides in length. In preferred embodiments, shRNAs have 5'-phosphate and 3'-hydroxyl groups.

Although the RNAi molecules useful according to the invention preferably contain nucleotide sequences that are fully complementary to a portion of the target locus, 100% sequence complementarity between the RNAi probe and the target locus is not required to practice the invention.

RNA molecules useful for RNAi may be chemically synthesized, for example using appropriately protected ribonucleoside phosphoramidites and a conventional DNA/RNA synthesizer. RNAs produced by such methodologies tend to be highly pure and to anneal efficiently to form siRNA duplexes or shRNA hairpin stem-loop structures. Following chemical synthesis, single stranded RNA molecules are deprotected, annealed to form siRNAs or shRNAs, and purified (*e.g.*, by gel electrophoresis or HPLC).

Alternatively, standard procedures may used for *in vitro* transcription of RNA from DNA templates carrying RNA polymerase promoter sequences (*e.g.*, T7 or SP6 RNA polymerase promoter sequences). Efficient *in vitro* protocols for preparation of siRNAs using T7 RNA polymerase have been described (Donzé and Picard, *Nucleic Acids Res.* 2002; 30: e46; and Yu *et al.*, *Proc. Natl. Acad. Sci. USA* 2002; 99: 6047-6052). Similarly, an efficient *in vitro* protocol for preparation of shRNAs using T7

RNA polymerase has been described (Yu *et al., supra*). The sense and antisense transcripts may be synthesized in two independent reactions and annealed later, or may be synthesized simultaneously in a single reaction.

RNAi molecules may be formed within a cell by transcription of RNA from an expression construct introduced into the cell. For example, both a protocol and an expression construct for *in vivo* expression of siRNAs are described in Yu *et al., supra*. Similarly, protocols and expression constructs for *in vivo* expression of shRNAs have been described (Brummelkamp *et al., supra*; Sui *et al., supra*; Yu *et al., supra*; McManus *et al., supra*; Paul *et al., supra*).

The expression constructs for *in vivo* production of RNAi molecules comprise RNAi encoding sequences operably linked to elements necessary for the proper transcription of the RNAi encoding sequence(s), including promoter elements and transcription termination signals. Preferred promoters for use in such expression constructs include the polymerase-III HI-RNA promoter (see, *e.g.*, Brummelkamp *et al., supra*) and the U6 polymerase-III promoter (see, *e.g.*, Sui *et al., supra*; Paul, *et al. supra*; and Yu *et al., supra*). The RNAi expression constructs can further comprise vector sequences that facilitate the cloning of the expression constructs. Standard vectors that maybe used in practicing the current invention are known in the art (*e.g.*, pSilencer 2.0-U6 vector, Ambion Inc., Austin, TX).

Antisense Nucleic Acids. In a specific embodiment, to achieve inhibition of expression of a NPC1L1 gene, the nucleic acid molecules of the invention can be used to design antisense oligonucleotides. An antisense oligonucleotide is typically 18 to 25 bases in length (but can be as short as 13 bases in length) and is designed to bind to a selected NPC1L1 mRNA. This binding prevents expression of that specific NPC1L1 protein. The antisense oligonucleotides of the invention comprise at least 6 nucleotides and preferably comprise from 6 to about 50 nucleotides. In specific aspects, the antisense oligonucleotides comprise at least 10 nucleotides, at least 15 nucleotides, at least 25, at least 30, at least 100 nucleotides, or at least 200 nucleotides.

The antisense nucleic acid oligonucleotides of the invention comprise sequences complementary to at least a portion of the corresponding NPC1L1 mRNA. However, 100% sequence complementarity is not required so long as formation of a stable duplex (for single stranded antisense oligonucleotides) or triplex (for double stranded antisense oligonucleotides) can be achieved. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense oligonucleotides. Generally, the longer the antisense oligonucleotide, the more base mismatches with the corresponding mRNA can be tolerated. One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

The antisense oligonucleotides can be DNA or RNA or chimeric mixtures, or derivatives or modified versions thereof, and can be single-stranded or double-stranded. The antisense oligonucleotides can be modified at the base moiety, sugar moiety, or phosphate backbone, or a combination thereof. For example, a NPC1L1-specific antisense oligonucleotide can comprise at least one modified base moiety selected from a group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

In another embodiment, the NPC1L1-specific antisense oligonucleotide comprises at least one modified sugar moiety, *e.g.*, a sugar moiety selected from arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the NPC1L1-specific antisense oligonucleotide comprises at least one modified phosphate backbone selected from a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

The antisense oligonucleotide can include other appending groups such as peptides, or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, *Proc. Natl. Acad. Sci. USA* 1989; 86: 6553-6556; Lemaitre *et al.*, *Proc. Natl. Acad. Sci. USA* 1987; 84: 648-652; PCT Publication No. WO 88/09810) or blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134), hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.*, *BioTechniques* 1988; 6: 958-976), intercalating agents (see, *e.g.*, Zon, *Pharm. Res.* 1988; 5: 539-549), etc.

In another embodiment, the antisense oligonucleotide can include α-anomeric oligonucleotides. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier *et al.*, *Nucl. Acids Res.* 1987; 15: 6625-6641).

In yet another embodiment, the antisense oligonucleotide can be a morpholino antisense oligonucleotide (*i.e.*, an oligonucleotide in which the bases are linked to 6-membered morpholine rings, which are connected to other morpholine-linked bases via non-ionic phosphorodiamidate intersubunit linkages). Morpholino oligonucleotides are resistant to nucleases and act by sterically blocking transcription of the target mRNA.

Similar to the above-described RNAi molecules, the antisense oligonucleotides of the invention can be synthesized by standard methods known in the art, *e.g.*, by use of an automated synthesizer. Antisense nucleic acid oligonucleotides of the invention can also be produced intracellularly by transcription from an exogenous sequence. For example, a vector can be introduced *in vivo* such that it is taken up by a cell within which the vector or a portion thereof is transcribed to produce an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, so long as it can be transcribed to produce the desired

antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. In another embodiment, "naked" antisense nucleic acids can be delivered to adherent cells via "scrape delivery", whereby the antisense oligonucleotide is added to a culture of adherent cells in a culture vessel, the cells are scraped from the walls of the culture vessel, and the scraped cells are transferred to another plate where they are allowed to re-adhere. Scraping the cells from the culture vessel walls serves to pull adhesion plaques from the cell membrane, generating small holes that allow the antisense oligonucleotides to enter the cytosol.

The present invention thus provides a method for inhibiting the expression of a NPC1L1 gene in a eukaryotic, preferably mammalian, and more preferably rat, mouse or human cell, comprising providing the cell with an effective amount of a NPC1L1-inhibiting antisenseoligonucleotide.

Ribozyme Inhibition. In another embodiment, the expression of NPC1L1 genes of the present invention can be inhibited by ribozymes designed based on the nucleotide sequence thereof. Ribozyme molecules catalytically cleave mRNA transcripts and can be used to prevent expression of the gene product. Ribozymes are enzymatic RNA molecules capable of catalyzing the sequence-specific cleavage of RNA (for a review, see Rossi, *Current Biology* 1994; 4: 469-471). The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage event. The composition of ribozyme molecules must include: (i) one or more sequences complementary to the target gene mRNA; and (ii) a catalytic sequence responsible for mRNA cleavage (see, *e.g.*, U.S. Patent No. 5,093,246).

According to the present invention, the use of hammerhead ribozymes is preferred. Hammerhead ribozymes cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The sole requirement is that the target mRNA has the following sequence of two bases: 5'-UG-3'. The construction of hammerhead ribozymes is known in the art, and described more fully in Myers, *Molecular Biology and Biotechnology: A Comprehensive Desk*

Reference, VCH Publishers, New York, 1995 (see especially Figure 4, page 833) and in Haseloff and Gerlach, *Nature* 1988; 334: 585-591.

Preferably, the ribozymes of the present invention are engineered so that the cleavage recognition site is located near the 5' end of the corresponding mRNA, *i.e.*, to increase efficiency and minimize the intracellular accumulation of non-functional mRNA transcripts.

As in the case of RNAi and antisense oligonucleotides, ribozymes of the invention can be composed of modified oligonucleotides (*e.g.*, for improved stability, targeting, etc.). These can be delivered to mammalian cells, and preferably mouse, rat, or human cells, which express the target NPC1L1 protein *in vivo*. A preferred method of delivery involves using a DNA construct "encoding" the ribozyme under the control of a strong constitutive pol III or pol II promoter, so that transfected cells will produce sufficient quantities of the ribozyme to destroy endogenous mRNA encoding the protein and inhibit translation. Because ribozymes, unlike antisense molecules, are catalytic, a lower intracellular concentration may be required to achieve an adequate level of efficacy.

Ribozymes can be prepared by any method known in the art for the synthesis of DNA and RNA molecules, as discussed above. Ribozyme technology is described further in *Intracellular Ribozyme Applications: Principals and Protocols*, Rossi and Couture eds., Horizon Scientific Press, 1999.

Triple Helix Formation. Nucleic acid molecules useful to inhibit NPC1L1 gene expression via triple helix formation are preferably composed of deoxynucleotides. The base composition of these oligonucleotides is typically designed to promote triple helix formation via Hoogsteen base pairing rules, which generally require sizeable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, resulting in TAT and CGC triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, *e.g.*, those containing a stretch of G

residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in GGC triplets across the three strands in the triplex.

Alternatively, sequences can be targeted for triple helix formation by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

Similarly to NPC1L1-specific RNAi, antisense oligonucleotides, and ribozymes, triple helix molecules of the invention can be prepared by any method known in the art. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides such as, *e.g.*, solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules can be generated by *in vitro* or *in vivo* transcription of DNA sequences "encoding" the particular RNA molecule. Such DNA sequences can be incorporated into a wide variety of vectors that incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters.

Other NPC1L1 Antagonists

NPC1L1 inhibitors also include small molecules inhibitors. For example, several NPC1L1 inhibitors have been identified and are set forth in Example 10. These inhibitors include, for example, 4-phenyl-4-piperidinecarbonitrile hydrochloride, 1-butyl-N-(2,6-dimethylphenyl)-2 piperidinecarboxamide, 1-(1-naphthylmethyl)piperazine, 3{1-[(2-methylphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3-[(4-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, 3-[(2-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, and N-(4-acetylphenyl)-2-thiophenecarboxamide, or derivatives thereof. Additional NPC1L1 antagonists, *e.g.*, small molecule antagonists, may be identified using, for example, the assays described herein.

Diagnostic Methods

A variety of methods can be employed for the diagnostic evaluation of lipid disorders, such as hyperlipidemia and other diseases and disorders associated with or mediated by NPC1L1, such as obesity, type II diabetes, cardiovascular disease, and stroke, and for the identification and evaluation of subjects experiencing or at risk for developing hyperlipidemia, *e.g.*, cholesterolemia and NPC1L1-associated conditions such as obesity, type II diabetes, cardiovascular disease, and stroke. These methods may also be employed for the diagnostic evaluation of diseases and disorders associated with decreased NPC1L1 such as anorexia, cachexia, and wasting.

These methods may utilize reagents such as the polynucleotide molecules and oligonucleotides of the present invention. The methods may alternatively utilize a NPC1L1 protein or a fragment thereof, or an antibody or antibody fragment that binds specifically to a NPC1L1 protein. Such reagents can be used for: (i) the detection of either an over- or an under-expression of the NPC1L1 gene relative to its expression in an unaffected state (*e.g.*, in a subject or individual not having a disease or disorder associated with or mediated by NPC1L1); or (ii) the detection of either an increase or a decrease in the level of the NPC1L1 protein relative to its level in an unaffected state; or (iii) the detection of an aberrant NPC1L1 gene product activity relative to the unaffected state; or (iv) the mislocalization of vesicular proteins such as caveolin or annexin.

In a preferred embodiment, a diagnostic method of the present invention utilizes quantitative hybridization (*e.g.*, quantitative *in situ* hybridization, Northern blot analysis or microarray hybridization) or quantitative PCR (*e.g.*, TaqMan®) using a NPC1L1-specific nucleic acid of the invention as a hybridization probe and PCR primers, respectively.

The present invention also provides a method for detecting cells which may have altered lipid or glucose metabolism in a test cell subjected to a treatment or stimulus or suspected of having been subjected to a treatment or stimulus, said method comprising:

(a) determining the expression level in the test cell of a nucleic acid molecule encoding a NPC1L1 protein; and

(b) comparing the expression level of the NPC1L1-encoding nucleic acid molecule in the test cell to the expression level of the same nucleic acid molecule in a control cell not subjected to a treatment or stimulus;

wherein a detectable change in the expression level of the NPC1L1-encoding nucleic acid molecule in the test cell compared to the expression level of the NPC1L1-encoding nucleic acid molecule in the control cell indicates that the test cell may have altered lipid or glucose metabolism .

According to the present invention, the detectable change in the expression level is any statistically significant change and preferably at least a 1.5-fold change as measured by any available technique such as hybridization or quantitative PCR (*see* the Definitions Section, above).

The test and control cells are preferably the same type of cells from the same species and tissue, and can be any cells useful for conducting this type of assay where a meaningful result can be obtained. Any cell type in which a NPC1L1-encoding nucleic acid molecule is ordinarily expressed, or in which a NPC1L1-encoding nucleic acid is expressed in connection with a treatment or stimulus affecting lipid or glucose metabolism may be used. For example, the test cell can be any cell derived from a tissue of an organism experiencing hyperlipidemia or another disease or disorder associated with or mediated by NPC1L1. Alternatively, the test cell can be any cell grown *in vitro* under specific conditions. When the test cell is derived from a tissue of an organism experiencing hyperlipidemia or another disease or disorder associated with or mediated by NPC1L1, it may or may not be known to be located in the region associated with disorder.

In one embodiment, the test and control cells are cells from the gastrontestinal system. Preferably, the test and control cells are enterocyte cells from the epithelium of the small intestine. The test and control cells can be derived from any appropriate organism, but are preferably human or mouse cells. In a specific embodiment, the test and control cells are from an animal model of lipid pathogenesis (*e.g.*, a mouse model

of hyperlipidemia) or any related disorder (*e.g.*, obesity, cardiovascular disease, or diabetes) and may or may not be isolated from that animal model. In another embodiment, the first cell is from a subject, such as a human or companion animal, for which the test is being conducted to determine the state of lipid or glucose metabolism that subject, and the second cell is an appropriate control cell. The first cell may or may not be isolated from the subject being tested. Both the test cell and the control cell must have the ability to express NPC1L1.

The control cell can be any cell which is known to have not been subjected to any treatment or stimulus associated with lipid or glucose metabolism. Preferably, the control cell is otherwise similar and treated identically to the test cell. For example, when the test cell is derived from a tissue of an animal experiencing hyperlipidemia or another disease or disorder associated with or mediated by NPC1L1, the control cell can be derived from an identical tissue or body part of a different animal from, preferably, the same species (or, alternatively, a closely related species) which animal is not experiencing hyperlipidemia or another disease or disorder associated with or mediated by NPC1L1. Alternatively, the control cell can be derived from an identical tissue or body part of the same animal from which the test cells are derived. However if this is the case, it should be established that the identical tissue or body part has not been subjected to any treatment or stimulus associated with lipid or glucose metabolism within the timeframe of the experiment. When the test cell is a cell grown *in vitro* under specific conditions, the control cell can be a similar cell grown *in vitro* in identical conditions but in the absence of the treatment or stimulus.

In one embodiment, the test cell has been exposed to a treatment or stimulus that simulates or mimics a lipid-related condition prior to determining the expression level of the nucleic acid molecule encoding the NPC1L1 protein, and the control cell is useful as an appropriate comparator cell to allow a determination of whether or not the test cell is exhibiting a lipid response. For example, where the test cell has been exposed to a treatment or stimulus that is, or that simulates or mimics, hyperlipidemia or another disease or disorder associated with or mediated by NPC1L1, the control cell has not been exposed to such a treatment or stimulus. In another embodiment, the test cell has been exposed to a compound that is being tested to determine whether it

simulates or mimics hyperlipidemia or another disease or disorder associated with or mediated by NPC1L1.

In one embodiment, the nucleic acid molecule the expression of which is being determined according to this method encodes a mammalian NPC1L1 polypeptide. In a specific embodiment, the nucleic acid molecule encodes a mouse NPC1L1 polypeptide comprising the amino acid sequence of SEQ ID NO: 3.

In one embodiment, the expression level of the nucleic acid molecule in each of the test and control cells is determined by quantifying the amount of NPC1L1-encoding mRNA present in the two cells. In another embodiment, the expression level of the nucleic acid molecule in each of the test and control cells is determined by quantifying the amount of NPC1L1 protein present in each of the two cells. Where the test cell has a detectable change in the expression level of the NPC1L1-encoding nucleic acid molecule compared to the expression level of the NPC1L1-encoding nucleic acid molecule in the control cell, a lipid response in the test cell has been detected.

To assay levels of a NPC1L1-encoding nucleic acid in a sample, a variety of standard nucleic acid isolation and quantification methods can be employed. As specified above, in a preferred embodiment, a diagnostic method of the present invention utilizes quantitative hybridization (*e.g.*, quantitative *in situ* hybridization, Northern blot analysis or microarray hybridization) or quantitative PCR (*e.g.*, TaqMan®) using NPC1L1-specific nucleic acids of the invention as hybridization probes and PCR primers, respectively.

In PCR-based assays, gene expression can be measured after extraction of cellular mRNA and preparation of cDNA by reverse transcription (RT). A sequence within the cDNA can then be used as a template for a nucleic acid amplification reaction. Nucleic acid molecules of the present invention can be used to design NPC1L1-specific RT and PCR oligonucleotide primers (such as, *e.g.*, SEQ ID NOS: 4-7). Preferably, the oligonucleotide primers are at least about 9 to about 30 nucleotides in length. The amplification can be performed using, *e.g.*, radioactively labeled or fluorescently-labeled nucleotides, for detection. Alternatively, enough amplified

product may be made such that the product can be visualized simply by standard ethidium bromide or other staining methods.

A preferred PCR-based detection method of the present invention is quantitative real time PCR (*e.g.*, TaqMan® technology, Applied Biosystems, Foster City, CA). This method is based on the observation that there is a quantitative relationship between the amount of the starting target molecule and the amount of PCR product produced at any given cycle number. Real time PCR detects the accumulation of amplified product during the reaction by detecting a fluorescent signal produced proportionally during the amplification of a PCR product.

For more details on quantitative real time PCR, see Gibson *et al.*, *Genome Res.* 1996; 6: 995-1001; Heid *et al.*, *Genome Res.* 1996; 6: 986-994; Livak *et al.*, *PCR Methods Appl.* 1995; 4: 357-362; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 1991; 88: 7276-7280.

SYBR Green Dye PCR (Molecular Probes, Inc., Eugene, OR), competitive PCR as well as other quantitative PCR techniques can also be used to quantify NPC1L1 gene expression according to the present invention.

NPC1L1 gene expression detection assays of the invention can also be performed *in situ* (*e.g.*, directly upon sections of fixed or frozen tissue collected from a subject, thereby eliminating the need for nucleic acid purification). Nucleic acid molecules of the invention or portions thereof can be used as labeled probes or primers for such *in situ* procedures (see, *e.g.*, Nuovo, *PCR in situ Hybridization: Protocols And Application*, Raven Press, New York, 1992). Alternatively, if a sufficient quantity of the appropriate cells can be obtained, standard quantitative Northern analysis can be performed to determine the level of gene expression using the nucleic acid molecules of the invention or portions thereof as labeled probes.

For *in vitro* cell cultures or *in vivo* animal models, the diagnostic reagents of the invention can be used in screening assays as surrogates lipid condition to identify compounds that affect expression of the NPC1L1 gene. For example, probes for the mouse NPC1L1 gene can be used for diagnosing individuals suspected of having a

condition associated with abnormal lipid or glucose metabolism, and also for monitoring the effectiveness therapy used to treat such condition.

Various techniques can be used to measure the levels of NPC1L1 protein in a sample, including the use of anti-NPC1L1 antibodies or antibody fragments described above. For example, anti-NPC1L1 antibodies or antibody fragments can be used to screen test compounds to identify those compounds that can modulate NPC1L1 protein production. For example, anti-NPC1L1 antibodies or antibody fragments can be used to detect the presence of the NPC1L1 protein by, *e.g.*, immunofluorescence techniques employing a fluorescently labeled antibody coupled with light microscopic, flow cytometric or fluorimetric detection methods. Such techniques are particularly preferred for detecting the presence of the NPC1L1 protein on the surface of cells. In addition, protein isolation methods such as those described by Harlow and Lane (*Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988) can also be employed to measure the levels of NPC1L1 protein in a sample.

Antibodies or antigen-binding fragments thereof may also be employed histologically, *e.g.*, in immunofluorescence or immunoelectron microscopy techniques, for *in situ* detection of the NPC1L1 protein. *In situ* detection may be accomplished by, *e.g.*, removing a tissue sample from a patient and applying to the tissue sample a labeled antibody or antibody fragment of the present invention. This procedure can be used to detect both the presence of the NPC1L1 protein and its distribution in the tissue. Additionally, antibodies or antigen-binding fragments may be used to detect NPC1L1 protein in the serum of cells, tissues, or animals that produce NPC1L1 protein.

Screening Methods

The present invention further provides a method for identifying a lead compound useful for modulating the expression of a NPC1L1-encoding nucleic acid, said method comprising:

(a) contacting a first cell with a test compound for a time period sufficient to allow the cell to respond to said contact with the test compound;

(b) determining the expression level of a NPC1L1-encoding nucleic acid molecule in the cell prepared in step (a); and

(c) comparing the expression level of the NPC1L1-encoding nucleic acid molecule determined in step (b) to the expression level of the NPC1L1-encoding nucleic acid molecule in a second (control) cell that has not been contacted with the test compound;

wherein a detectable change in the expression level of the NPC1L1-encoding nucleic acid molecule in the first cell in response to contact with the test compound compared to the expression level of the NPC1L1-encoding nucleic acid molecule in the second (control) cell that has not been contacted with the test compound, indicates that the test compound modulates the expression of the NPC1L1-encoding nucleic acid and is a candidate compound for the treatment of a disorder associated with abnormal lipid or glucose metabolism.

In one embodiment, the candidate compound decreases the expression of the NPC1L1-encoding nucleic acid molecule. In another embodiment, the candidate compound increases the expression of the NPC1L1-encoding nucleic acid molecule. In another embodiment, the first and second cells are incubated under conditions that induce the expression of a NPC1L1-encoding nucleic acid molecule, but the test compound is tested for its ability to inhibit or reduce the induction of such expression in the first cell. In another embodiment, the first and second cells are incubated under conditions that induce the expression of a NPC1L1-encoding nucleic acid molecule, but the test compound is tested for its ability to potentiate the induction of such expression in the first cell.

The test compound can be, without limitation, a small organic or inorganic molecule, a polypeptide (including an antibody, antibody fragment, or other immunospecific molecule), an oligonucleotide molecule, a polynucleotide molecule, or a chimera or derivative thereof. Test compounds that specifically bind to a NPC1L1-encoding nucleic acid molecule or to a NPC1L1 protein of the present invention can be identified, for example, by high-throughput screening (HTS) assays, including cell-based and cell-free assays, directed against individual protein targets. Several methods

of automated assays that have been developed in recent years enable the screening of tens of thousands of compounds in a short period of time (see, e.g., U.S. Patent Nos. 5,585,277, 5,679,582, and 6,020,141). Such HTS methods are particularly preferred.

The first and second cells are preferably the same types of cells, and can be any cells useful for conducting this type of assay where a meaningful result can be obtained. Such cells can be prokaryotic, but are preferably eukaryotic. Such eukaryotic cells are preferably mammalian cells, and more preferably mouse or human cells. Both the first and second cell must have the ability to express NPC1L1. In one non-limiting embodiment, the first and second cells are cells that have been genetically modified to express or over-express a NPC1L1 nucleic acid molecule. In another non-limiting embodiment, the first and second cells are cells that express a NPC1L1 nucleic acid molecule, either naturally (*e.g.*, cells lining the small intestine) or in response to an appropriate stimulus. In one embodiment, the first and second cells have been exposed to a condition or stimulus that is, or that simulates or mimics, a lipid condition prior to, or at the same time as, exposing the cells to the test compound to determine the effect of the test compound on the expression level of the nucleic acid molecule encoding the NPC1L1 polypeptide.

In one embodiment, the first and second cells are from an animal model of a disease or disorder associated with or mediated by NPC1L1 (*e.g.*, mouse model of hypercholestolemia, obesity, diabetes, stroke or cardiovascular disease), and may or may not be isolated from that animal model. In another embodiment, the first cell is from a subject, such as a human or companion animal, and the second cell is an appropriate control cell. The first cell may or may not be isolated from the subject being tested.

In one embodiment, the nucleic acid molecule the expression of which is being determined according to this method encodes a mammalian NPC1L1 polypeptide. In a specific embodiment, the nucleic acid molecule encodes a mouse NPC1L1 polypeptide. In another embodiment, the mouse NPC1L1 polypeptide comprises the amino acid sequence of SEQ ID NO:3.

The expression level of the nucleic acid molecule in each of the first and second cells can be determined by quantifying and comparing the amount of NPC1L1-encoding mRNA present in each of the first and second cells. Alternatively, the expression level of the nucleic acid molecule in each of the first and second cells can be determined by quantifying and comparing the amount of NPC1L1 protein present in the first and second cells. Where the first cell has a detectable change in the expression level of the nucleic acid encoding a NPC1L1 protein compared to the expression level of the nucleic acid encoding the NPC1L1 protein in the second cell, the test compound is identified as a candidate compound useful for modulating the expression of a NPC1L1-encoding nucleic acid.

The present invention also provides a method for identifying a candidate compound that modulates an NPC1L1 polypeptide. In one embodiment, the present invention provides a method for identifying a ligand or other binding partner to the NPC1L1 protein of the present invention, which comprises bringing a labeled test compound in contact with the NPC1L1 protein or a fragment thereof and measuring the amount of the labeled test compound bound to the NPC1L1 protein or to the fragment thereof.

In another embodiment, the present invention provides a method for identifying a ligand or other binding partner to the NPC1L1 protein of the present invention, which comprises bringing a labeled test compound in contact with cells or cell membrane fraction containing the NPC1L1 protein, and measuring the amount of the labeled test compound bound to the cells or the membrane fraction.

In yet a third embodiment, the present invention provides a method for identifying a ligand or other binding partner to the NPC1L1 polypeptide of the present invention, which comprises culturing a transfected cell containing the DNA encoding the NPC1L1 protein under conditions that permit or induce expression of the NPC1L1 protein, bringing a labeled test compound in contact with the NPC1L1 protein expressed on a membrane of said cell, and measuring the amount of the labeled test compound bound to the NPC1L1 protein.

For example, the ligand or binding partner of the NPC1L1 protein of the present invention can be determined by the following procedures. First, a standard NPC1L1 preparation can be prepared by suspending cells or membranes containing the NPC1L1 protein in a buffer appropriate for use in the determination method. Any buffer can be used so long as it does not inhibit the ligand-NPC1L1 binding. Such buffers include, e.g., a phosphate buffer or a Tris-HCl buffer having pH of 4 to 10 (preferably pH of 6 to 8). For the purpose of minimizing non-specific binding, a surfactant such as CHAPS, Tween-80TM (manufactured by Kao-Atlas Inc.), digitonin or deoxycholate, and various proteins such as bovine serum albumin or gelatin, may optionally be added to the buffer. For the purpose of suppressing degradation of the NPC1L1 or ligand by proteases, a protease inhibitor such as PMSF, leupeptin, E-64 (manufactured by Peptide Institute, Inc.) and pepstatin can be added. A given amount (e.g., 5,000 to 500,000 cpm) of the test compound labeled with [^{3}H], [^{125}I], [^{14}C], [^{35}S] or the like can be added to about 0.01 ml to 10 ml of the solution containing NPC1L1. To determine the amount of non-specific binding (NSB), a reaction tube containing an unlabeled test compound in a large excess is also prepared. The reaction is carried out at about 0 to 50°C, preferably about 4 to 37oC for about 20 minutes to about 24 hours, preferably about 30 minutes to about 3 hours. After completion of the reaction, the cells or membranes containing any bound ligand are separated, e.g., the reaction mixture is filtered through glass fiber filter paper and washed with an appropriate volume of the same buffer. The residual radioactivity on the glass fiber filter paper can be measured by means of a liquid scintillation counter or λ-counter. A test compound exceeding 0 cpm obtained by subtracting NSB from the total binding (B) (B minus NSB) may be selected as a ligand or binding partner of the NPC1L1 protein of the present invention.

Additionally, any of a variety of known methods for detecting protein-protein interactions may also be used to detect and/or identify proteins that bind to a NPC1L1 gene product. For example, co-immunoprecipitation, chemical cross-linking and yeast two-hybrid systems as well as other techniques known in the art may be employed. As an example in a yeast two-hybrid assay, a host cell harbors a construct that expresses a NPC1L1 protein or fragment thereof fused to a DNA binding domain and another construct that expresses a potential binding-partner fused to an activation domain. The host cell also includes a reporter gene that is expressed in response to binding of the

NPC1L1 protein-partner complex (formed as a result of binding of binding-partner to the NPC1L1 protein) to an expression control sequence operatively associated with the reporter gene. Reporter genes for use in the yeast two-hybrid assay of the invention encode detectable proteins, including, but by no means limited to, chloramphenicol transferase (CAT), β galactosidase (β gal), luciferase, green fluorescent protein (GFP), alkaline phosphatase, and other genes that can be detected, e.g., immunologically (by antibody assay). See the Mammalian MATCHMAKER Two-Hybrid Assay Kit User Manual from Clontech (Palo Alto, CA) for further details on mammalian two-hybrid methods.

All of the screening methods described herein can be modified for use in high-throughput screening, *e.g.*, using microarrays.

Microarrays

Protein arrays. Protein arrays are solid-phase, ligand binding assay systems using immobilized proteins on surfaces that are selected from glass, membranes, microtiter wells, mass spectrometer plates, and beads or other particles. The ligand binding assays using these arrays are highly parallel and often miniaturized. Their advantages are that they are rapid, can be automated, are capable of high sensitivity, are economical in their use of reagents, and provide an abundance of data from a single experiment.

Automated multi-well formats are the best-developed HTS systems. Automated 96-well plate-based screening systems are the most widely used. The current trend in plate based screening systems is to reduce the volume of the reaction wells further, thereby increasing the density of the wells per plate (96 wells to 384 wells, and 1,536 wells per plate). The reduction in reaction volumes results in increased throughput, dramatically decreased bioreagent costs, and a decrease in the number of plates that need to be managed by automation. For a description of protein arrays that can be used for HTS, see, e.g., U.S. Patents No. 6,475,809; 6,406,921; and 6,197,599; and International Publications No. WO 00/04389 and WO 00/07024.

For construction of arrays, sources of proteins include cell-based expression systems for recombinant proteins, purification from natural sources, production in vitro by cell-free translation systems, and synthetic methods for peptides. For capture arrays

and protein function analysis, it is important that proteins are correctly folded and functional. This is not always the case, e.g., where recombinant proteins are extracted from bacteria under denaturing conditions, whereas other methods (isolation of natural proteins, cell free synthesis) generally retain functionality. However, arrays of denatured proteins can still be useful in screening antibodies for cross-reactivity, identifying auto-antibodies, and selecting ligand binding proteins.

The immobilization method used should be reproducible, applicable to proteins of different properties (size, hydrophilic, hydrophobic), amenable to high throughput and automation, and compatible with retention of fully functional protein activity. Both covalent and non-covalent methods of protein immobilization can be used. Substrates for covalent attachment include, e.g., glass slides coated with amino- or aldehyde-containing silane reagents (Telechem). In the Versalinx™ system (Prolinx), reversible covalent coupling is achieved by interaction between the protein derivatized with phenyldiboronic acid, and salicylhydroxamic acid immobilized on the support surface. Covalent coupling methods providing a stable linkage can be applied to a range of proteins. Non-covalent binding of unmodified protein occurs within porous structures such as HydroGel™ (PerkinElmer), based on a 3-dimensional polyacrylamide gel.

Cell-Based Arrays. Cell-based arrays combine the technique of cell culture in conjunction with the use of fluidic devices for measurement of cell response to test compounds in a sample of interest, screening of samples for identifying molecules that induce a desired effect in cultured cells, and selection and identification of cell populations with novel and desired characteristics. High-throughput screens (HTS) can be performed on fixed cells using fluorescent-labeled antibodies, biological ligands and/or nucleic acid hybridization probes, or on live cells using multicolor fluorescent indicators and biosensors. The choice of fixed or live cell screens depends on the specific cell-based assay required.

There are numerous single- and multi-cell-based array techniques known in the art. Recently developed techniques such as micro-patterned arrays (described, e.g., in International PCT Publications WO 97/45730 and WO 98/38490) and microfluidic arrays provide valuable tools for comparative cell-based analysis. Transfected cell microarrays are a complementary technique in which array features comprise clusters of cells overexpressing defined cDNAs. Complementary DNAs cloned in expression

vectors are printed on microscope slides, which become living arrays after the addition of a lipid transfection reagent and adherent mammalian cells (Bailey et al., Drug Discov. Today 2002; 7(18 Suppl): S113-8). Cell-based arrays are described in detail in, e.g., Beske, Drug Discov. Today 2002; 7(18 Suppl): S131-5; Sundberg et al., Curr. Opin. Biotechnol. 2000; 11: 47-53; Johnston et al., Drug Discov. Today 2002; 7: 353-63; U.S. Patents No. 6,406,840 and 6,103,479, and U.S. published patent application No. 2002/0197656. For cell-based assays specifically used to screen for modulators of ligand-gated ion channels, see Mattheakis et al., Curr. Opin. Drug Discov. Devel. 2001; 1: 124-34; and Baxter et al., J. Biomol. Screen. 2002; 7: 79-85.

For detection of molecules using screening assays, a molecule (e.g., an antibody or polynucleotide probe) can be detectably labeled with an atom (e.g., radionuclide), detectable molecule (e.g., fluorescein), or complex that, due to its physical or chemical property, serves to indicate the presence of the molecule. A molecule can also be detectably labeled when it is covalently bound to a "reporter" molecule (e.g., a biomolecule such as an enzyme) that acts on a substrate to produce a detectable product. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Labels useful in the present invention include, but are not limited to, biotin for staining with labeled avidin or streptavidin conjugate, magnetic beads (e.g., Dynabeads™), fluorescent dyes (e.g., fluorescein, fluorescein-isothiocyanate (FITC), Texas red, rhodamine, green fluorescent protein, enhanced green fluorescent protein, lissamine, phycoerythrin, Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX [Amersham], SyBR Green I & II [Molecular Probes], and the like), radiolabels (e.g., 3H, 125I, 35S, 14C, or 32P), enzymes (e.g., hydrolases, particularly phosphatases such as alkaline phosphatase, esterases and glycosidases, or oxidoreductases, particularly peroxidases such as horse radish peroxidase, and the like), substrates, cofactors, inhibitors, chemiluminescent groups, chromogenic agents, and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Examples of patents describing the use of such labels include U.S. Patents No. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

Means of detecting such labels are known to those of skill in the art. For example, radiolabels and chemiluminescent labels can be detected using photographic

film or scintillation counters; fluorescent markers can be detected using a photo-detector to detect emitted light (e.g., as in fluorescence-activated cell sorting); and enzymatic labels can be detected by providing the enzyme with a substrate and detecting, e.g., a colored reaction product produced by the action of the enzyme on the substrate.

Activity Assays

The present invention further provides a method for studying additional biological activities of the NPC1L1 protein. The biological activity of the NPC1L1 protein can be studied using intact cells that express the NPC1L1 protein (either naturally, e.g., as a result of a stimulus or treatment, or heterologously), membrane fractions comprising the NPC1L1 protein, the isolated NPC1L1 protein, soluble NPC1L1 fragments, or NPC1L1 fusion proteins. For example, a biological activity of the NPC1L1 protein can be studied by measuring in a cell that heterologously expresses the NPC1L1 protein the activities that promote or suppress the production of an "index substance", change in cell membrane potential, phosphorylation of intracellular proteins, activation of c-fos, pH reduction, etc.

NPC1L1-mediated activities can be determined by any known method. For example, cells containing the NPC1L1 protein can first be cultured on a multi-well plate, etc. Prior to the activity determination, the medium can be replaced with fresh medium or with an appropriate non-cytotoxic buffer, followed by incubation for a given period of time in the presence of a test compound, etc. Subsequently, the cells can be extracted or the supernatant can be recovered and the resulting product can be quantified by appropriate procedures. Where it is difficult to detect the production of the "index substance" for the cell-stimulating activity due to a degrading enzyme contained in the cells, an inhibitor against such a degrading enzyme may be added prior to the assay. For detecting activities such as the cAMP production suppression activity, the baseline production in the cells is increased by forskolin or the like and the suppressing effect on the increased baseline.

Methods of Treatment

The present invention provides methods for treating, *e.g.*, ameliorating, preventing, inhibiting, reducing the symptoms of, or delaying a condition that can be treated by modulating expression of a NPC1L1-encoding nucleic acid molecule or a NPC1L1 protein, comprising administering to a subject in need of such treatment a therapeutically effective amount of a compound that modulates expression of a NPC1L1-encoding nucleic acid molecule or a NPC1L1 protein.

Conditions that can be treated or prevented using the methods disclosed herein include those in which there are abnormalities in regulating lipid metabolism or responses, including cellular influx or efflux, endocytosis, or intracellular trafficking, transport, or localization of lipids, *e.g.*, cholesterol, fatty acids, triglycerides, and sphingolipids. Such conditions include those that are associated with hyperlipidemia, including diet-induced hypercholesterolemia, obesity, cardiovascular disease, and stroke. In addition, conditions associated with aberrant glucose metabolism and transport, *e.g.*, diabetes (*e.g.*, type II diabetes) can also be treated using the methods disclosed herein. Furthermore, conditions associated with decreased NPC1L1 expression or activity, such as anorexia, cachexia, and wasting, may also be treated or prevented using the methods disclosed herein.

The term "therapeutically effective amount" is used here to refer to: (i) an amount or dose of a compound sufficient to detectably change the level of expression of a NPC1L1-encoding nucleic acid in a subject; or (ii) an amount or dose of a compound sufficient to detectably change the level of activity of a NPC1L1 protein in a subject; or (iii) an amount or dose of a compound sufficient to cause a detectable improvement in a clinically significant symptom or condition (e.g., amelioration of hypercholesterolemia) in a subject.

In a preferred embodiment, the therapeutically effective amount of a compound reduces or inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

Formulations and Administration

A candidate compound useful in conducting a therapeutic method of the present invention is advantageously formulated in a pharmaceutical composition with a pharmaceutically acceptable carrier. The candidate compound may be designated as an active ingredient or therapeutic agent for the treatment of dietary hypercholesterolemia or other disorder involving lipid or glucose metabolism or transport.

The concentration of the active ingredient depends on the desired dosage and administration regimen, as discussed below. Suitable dose ranges of the active ingredient are from about 0.01 mg/kg to about 1500 mg/kg of body weight per day.

Therapeutically effective compounds can be provided to the patient in standard formulations, and may include any pharmaceutically acceptable additives, such as excipients, lubricants, diluents, flavorants, colorants, buffers, and disintegrants. The formulation may be produced in useful dosage units for administration by oral, parenteral, transmucosal, intranasal, rectal, vaginal, or transdermal routes. Parental routes include intravenous, intra-arteriole, intramuscular, intradermal, subcutaneous, intraperitoneal, intraventricular, intrathecal, and intracranial administration.

The pharmaceutical composition may also include other biologically active substances in combination with the candidate compound. Such substances include but are not limited to lovastin and ezetimibe.

The pharmaceutical composition can be added to a retained physiological fluid such as blood or synovial fluid. For CNS administration, a variety of techniques are available for promoting transfer of the therapeutic agent across the blood brain barrier, including disruption by surgery or injection, co-administration of a drug that transiently opens adhesion contacts between CNS vasculature endothelial cells, and co-administration of a substance that facilitates translocation through such cells.

In another embodiment, the active ingredient can be delivered in a vesicle, particularly a liposome.

In another embodiment, the therapeutic agent can be delivered in a controlled release manner. For example, a therapeutic agent can be administered using

intravenous infusion with a continuous pump, in a polymer matrix such as poly-lactic/glutamic acid (PLGA), in a pellet containing a mixture of cholesterol and the active ingredient (SilasticR™; Dow Corning, Midland, MI; see U.S. Patent No. 5,554,601), by subcutaneous implantation, or by transdermal patch

EXAMPLES

The present invention is further described by way of the following particular examples. However, the use of such examples is illustrative only and is not intended to limit the scope or meaning of this invention or of any exemplified term. Nor is the invention limited to any particular preferred embodiment(s) described herein. Indeed, many modifications and variations of the invention will be apparent to those skilled in the art upon reading this specification, and such "equivalents" can be made without departing from the invention in spirit or scope. The invention is therefore limited only by the terms of the appended claims, along with the full scope of equivalents to which the claims are entitled.

EXAMPLE 1: Intracellular Localization of the NPC1L1 Protein

Previous studies have revealed localization of NPC1 to the late endosome compartment of cells. The presence of NPC1 in this critical sorting region is consistent with the molecular etiology of Niemann-Pick C1 disease, which includes disruptions of cholesterol trafficking, storage, and secretion. Whether the NPC1L1 of the present invention localizes to the same region, however, is unclear. Although NPC1 and NPC1L1 have a number of common structural and functional domains, they also have different targeting sequences, suggesting distinct patterns of localization in the cell. In addition, another group has suggested that NPC1L1 molecule is present on the plasma membrane of enterocytes lining the small-intestine, a location consistent with their proposal that NPC1L1 is a transporter of dietary cholesterol and target of the anti-cholesterol drug ezitimibe. However, a recent study by Smart el al. (PNAS (2004) 101:345-3455, which presents evidence in both zebrafish and mouse systems that the target of ezitimibe is an annexin--caveolin heterocomplex, which is implicated as key

mediator in the intestinal transport and trafficking of cholesterol. The present invention addresses this issue with a set of reagents and approaches to determine NPC1L1 localization.

Methods

Production and purification of NPC1L1 antigen. A specific fragment of human NPC1L1 was amplified by PCR using the primers:

5'-GCGGGATCCGAACCGGTCCAGCTACAGGTA-3' (SEQ ID NO: 4) and

5'-GCGGAATTCCTCGAGGATGGGCAGGTCTTCAG-3' (SEQ ID NO: 5) spanning nucleotides 1302-1961 of SEQ ID NO: 2 and amino acids 416-635 of SEQ ID NO: 3. The amplified fragment was inserted into the pET-TRX expression vector, and the resulting recombinant plasmid was introduced into the host cell line, *E. coli* Bl21 (DE3) plysS. Purified NPC1L1 polypeptide was obtained by induced expression of the transformed cells followed by nickel affinity chromatography on a BioCAD system (Perseptive Biosystems, Framingham, MA).

Production and purification of anti-NPC1L1 antibodies. The NPC1L1 polypeptide was injected into two rabbits and polyclonal antisera was subsequently collected. Antiserum was sequentially purified in two affinity chromatography steps: (i) removal of Trx antibodies on a Trx-Affigel 10 column (BioRad, Hercules, CA); and (ii) purification of IgG antibodies on a Protein A-Sepharose column (Amersham Biosciences, Piscataway, NJ).

Construction of NPC1L1 fusion vectors and RFP-reporter constructs. Monomeric (m) YFP and CFP were generated using eYFP and eCFP plasmids (Clontech) as templates. The L221K and Q69M mutations for mYFP and the L221K mutation in mCFP were created using the megaprimer PCR mutagenesis method and verified by sequencing. To generate mYFP and mCFP fusions with NPC1L1, the stop codon of the human NPC1L1 sequence (GenBank accession number AY515256 was removed by PCR amplification and the resulting cDNA was verified by sequencing and fused to the mYFP and mCFP cDNAs. To introduce a Flag tag into NPC1L1, an

adapter encoding the Flag tag amino acid sequence DYKDDDDK (SEQ ID NO: 29) was ligated in frame into the NPC1L1 at the unique *BsmI* restriction site. To generate a construct of RFP driven by the human ABCA1 promoter the genomic sequence of the promoter was amplified (nucleotides -189 to +32) and inserted into the pDsRed-Express vector (Clontech).

Tissue culture, transfection, and immunofluorescence studies. All cells, including COS7, NT2 and Caco-2 cells, obtained from ATCC (Manassas, VA), were grown in DMEM supplemented with 2mM glutamine, 10% FCS and Gentamicin at 37°C and 5% CO_2 in a humidified incubator. Cells were transfected using 4ul Lipofectamine and 6 µl Plus reagent (Invitrogen, Carlsbad. CA), according to the manufacturer's recommendations. At 24 hr post-transfection the cells were either viewed live or they were fixed with ice-cold methanol at 4°C for 6 min. Cells were processed for immunofluorescence using standard procedures and 1 µg/ml of rabbit polyclonal antibody or 2 ug/ml of M2 anti-Flag antibody (Sigma, St. Louis, MO), followed by a 1:1000 dilution of the appropriate secondary antibody, either goat anti-rabbit IgG-Alexa 488 (Molecular Probes, Eugene, OR) or sheep anti-mouse IgG-FITC (Jackson Immunoresearch Laboratories, West Grove, PA). Cells were mounted in Fluoromount-G (Southern Biotechnology Associates, Birmingham, AL) and photographed using a Nikon Eclipse microscope equipped with a CCD camera.

Plasma membrane labeling assay. COS7 cells transfected with either Flag-tagged NPC1L1 or CD32 were labeled for 1 hr at 37°C with 100 µCi ^{S35}-Met/^{S35}-Cys in cell medium deficient in these amino acids. Following a 2hr chase period in DMEM complete medium, cells were removed from dishes using PBS containing 1 mM EDTA, washed in PBS and split equally into two eppendorfs. 2 µg of anti-Flag or anti-CD32 antibodies were added to half the samples and incubated on a rotating mixer at 4°C for 30 min. Cells were washed twice with cold PBS and all samples were lysed in 500 µl lysis buffer (NPC1L1: 100mM sodium phosphate pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% igepal, 0.01% SDS; CD32: 50mM Tris pH 7.4, 120 mM NaCl, 25 mM KCl, 0.2% Triton X100) containing proteinase inhibitor cocktail for 1hr 30 min at 4°C. Lysates were cleared by centrifugation at 20,000 g for 10 min at 4°C. Samples previously incubated with antibody were transferred to tubes containing 20 µl protein

G-agarose beads (Roche Applied Science, Indianapolis, IN) and incubated overnight at 4°C. Remaining samples were incubated at 4°C for 1hr with 3 µg anti-Flag/anti-CD32 antibodies, after which they were transferred to tubes containing protein G-agarose and incubated overnight at 4°C. Samples were washed four times in CD32 lysis buffer and once in NET1 buffer (50mM Tris pH 7.4, 0.5M NaCl, 1mM EDTA, 0.1% igepal, 0.25% gelatin, 0.02% sodium azide) and electrophoresed on a 4-20% bis-tris NUPAGE gel (Invitrogen, Carlsbad, CA) using the MOPS buffer system, until adequate separation was achieved. Gels were fixed in a solution of 10% acetic acid, 20% methanol for 10 min and soaked in Amplify solution for 15 min, before drying and exposing to film.

Results

In one set of experiments, the purified anti-NPC1L1 polyclonal antibodies were used to determine the in situ localization of endogenous NPC1L1 in the human NT2 cell line. As visualized by indirect immunofluorescence, endogenous NPC1L1 showed a perinuclear, ER to Golgi distribution (Figure 1a). Colocalization studies with various subcellular organelle markers (data not shown) confirmed the presence of NPC1L1 in the ER and Golgi. Notably, endogenous NPC1L1 was not present in the late endosomal/lysosomal compartment -- in sharp contrast to the previously established residence of NPC1 in late endosomes (Higgins et al., (1999) Mol. Genet. Metab. 68: 1-13).

In another experiment, COS7 cells were visualized by fluorescent microscopy, following transient transfection of the expression vector comprising NPC1L1 fused to the Flag epitope. Consistent with the NT2 studies, the NPC1L1-flag fusion protein also localized predominantly to the ER and Golgi (Figure 1b).

In addition, live Caco-2 cells were visualized by fluorescent microscopy, following transient transfection of the expression vector comprising NPC1L1 fused to mYFP. Again, the results reveal predominant localization of the NPC1L1 fusion to the ER and Golgi (Figure 1c).

In addition, colocalization experiments (shown in Davies et al., *J Biol Chem.* 2005) revealed that NPC1L1 localizes in an intracellular vesicular compartment with the marker protein Rab5.

In a final experiment, the membrane labeling assay was used as a sensitive detection method to confirm the intracellular localization of NPC1L1. In accord with the other findings, very little NPC1L1 can be labeled on the plasma membrane.

EXAMPLE 2: NPC1L1 mRNA Expression in Human and Mouse Tissues

Methods

Real time PCR quantitation. Human and mouse multiple tissue cDNA panels that had been normalized to four different control genes by the manufacturer (BD Biosciences Clontech, Palo Alto, CA) were amplified to detect only the full-length form of NPC1L1. Real-time PCR amplification was achieved using the Lightcycler 2 (Roche Applied Sciences). Data analysis was carried out using the accompanying software (v. 4.0). The primers used for amplifying mouse NPC1L1 were: 5'-GCTTCTTCCGCAAGATATACACTCCC-3' (SEQ ID NO: 6) and 5'-GAGGATGCAGCAATAGC CACATAAGAC-3' (SEQ ID NO: 7). The primers used for human NPC1L1 were 5'-TATCTTCCCTGGTTCCTGAACGAC-3' (SEQ ID NO: 8) and 5'-CCGCAGAGCTTCTGTGTAATCC-3' (SEQ ID NO: 9). For both the amplification cycles used were 95°C for 10 sec, 58°C for 20 sec and 72°C for 20 sec. Relative quantitation was carried out using external standards and a linear fit method and each sample was amplified in three separate experiments. All statistical calculations were obtained using Microsoft Excel.

Results

To further the functional studies of NPC1L1 the distribution of NPC1L1 mRNA expression was examined in both human and mouse tissues. In human tissues NPC1L1 is predominantly expressed in liver with detectable levels in lung, heart, brain, pancreas and kidney, ranging in expression from about 0.5 to 3% of liver expression (Figure 2). Since it has been reported that mouse NPC1L1 is predominantly expressed in the small intestine (Higgins et al., 2001), analyses using a human panel of

digestive tract tissues were also carried out. Human NPC1L1 is expressed in the small intestine at 1-4% of the levels expressed in liver (Figure 2a-c) suggesting that there are significant differences between the expression of human and mouse NPC1L1. Interestingly, analyses of mouse tissues suggests a predominant role for NPC1L1 in embryogenesis since its highest expression is found in 17-day embryos; low but detectable expression was found in lung, heart, spleen and kidney and elevated expression in brain, muscle and testis (Figure 2a-c).

EXAMPLE 3: Lipid Uptake Function of NPC1L1 Function

Introduction

NPC1L1 and NPC1 share a number of key structural features, including thirteen membrane spanning regions and a putative sterol sensitive motif. Accordingly, an important question is whether NPC1L1 shares some of the same functional properties as NPC1L1, specifically in the transport and movement of lipids. The present invention addresses the issue with respect to assays in bacterial cells.

Methods

E. coli fatty acid transport assays. The predicted signal peptide of human NPC1L1, amino acids 1-33, was removed and the remaining full-length sequence, encoding amino-acids 33-1359, was cloned in-frame with the amino-terminal *E. coli* Omp A signal peptide sequence in the vector pIN III OmpA, as previously described for NPC1 (Davies et al., 2000). NPC1L1 was then expressed in the 2.1.1 strain of *E.coli*, as previously described (Davies et al., 2000) Briefly, *E.coli* cultures grown to log phase were induced to express NPC1L1 using 1mM IPTG and grown for 1-2 hours. They were then diluted to an OD600 of 0.1 and incubated at 37°C for 5-15 min in saline containing 0.1M TRIS, Ph7.5, 1nM ^{3}H sodium oleate and 105 nM cold sodium oleate. Cell pellets were resuspended in water and ^{3}H sodium oleate was quantitated by scintillation counting.

Results

NPC1L1 was expressed in an engineered *E. coli* strain, designed for lipid transport studies (Davies et al., 2000). *E. coli* cells exhibited an increase in fatty acid accumulation compared to cells harboring a vector control (Figure 3), albeit at a lower level than cells expressing NPC1 "indicating that NPC1L1 might have a function similar to that of NPC1 in a different intracellular location. These and other data (Davies et al., J Biol Chem 2005) indicate that NPC1L1 is a Rab5 colocalized intracellular protein that appears to share lipid permease activity with NPC1." .

EXAMPLE 4: Generation of NPC1L1 Knockout Mice

Introduction

Unlike NPC1, no human disease arising from mutations in NPC1L1 is currently known. To address this issue, the present invention discloses the isolation of the mouse NPC1L1 gene and its targeted disruption in the appropriate mouse strain. In this regard, the C57BL6 strain was chosen, given its established utility in the study of cholesterol-related diseases, including atherosclerosis.

Methods

Isolation of mouse NPC1L1 gene. The genomic databases for BACs containing the mouse genomic sequence were searched and one clone that contained the mouse NPC1L1 promoter and entire coding region was identified. This clone, BAC RP23 64P22, accession number AC079435, from a C57BL6/J female mouse library, was obtained from BacPac Resources, Children's Hospital Oakland Research Institute (Oakland, CA). DNA was isolated using a BAC DNA isolation kit, as recommended (InCyte Genomics, St Louis, MO).

The mouse genomic nucleic acid sequence is provided in SEQ ID NO: 1. (The human genomic sequence is also provided in SEQ ID NO: 20. The NPC1L1 human cDNA is also presented in SEQ ID NO: 21 (GenBank Accession No. NM_013389), and corresponding amino acid in SEQ ID NO: 22 (GenBank Accession No.: NP_037521).

Targeted disruption of the endogenous NPC1L1 locus. A pGem7zf+ (Promega)-based construct was engineered to contain nucleotides 84689 to 96003 of the mouse NPC1L1 gene (accession number AC079435), spanning the promoter region to intron 6. The gene was disrupted at the unique *Afe I* restriction enzyme site in exon 2 of the mouse NPC1L1 sequence (at 91263) by insertion of phosphoglycerate kinase neomycin phosphotransferase hybrid gene (PGK-neo), in an antisense direction. This disrupts the coding sequence after cDNA nucleotide 601 so that no more than 200 amino acids of NPC1L1 can be expressed. Thus the expression of all alternatively spliced forms of the gene is abrogated. Homologous recombination and selection for neomycin resistant knockout clones using C57BL6 ES cells (Taconic, Germantown, NY) was carried out by Cell and Molecular Technologies (Phillipsburg, NJ).

About 150 neo-resistant ES clones were obtained, 4 of which were correctly targeted by homologous recombination of the neomycin cassette into the NPC1L1 gene, clones 13, 19, 44 and 144. These were identified by PCR screening using two sets of primers, each containing one primer outside the NPC1L1 targeting cassette and one within the neomycin gene hybrid. At the 5' end, these were 5'-CCTCCCTATTCCCCAAGATGTATGC -3' (SEQ ID NO: 10) in the NPC1L1 gene at 83538 and 5'-GGAGAGGCTATTCGGCTATGAC-3' (SEQ ID NO: 11) in the neomycin cassette. At the 3' end these were: 5'-CTGGGCTCCCTCTTAGAATAACCTA-3' SEQ ID NO: 12) at 96815 and 5'-GGAGAGGCTATTCGGCTATGAC-5' (SEQ ID NO: 13) in the neomycin cassette. Long-range amplifications were achieved using the Failsafe PCR system (Epicentre, Madison, WI) with buffer F and 30 cycles of: 94°C for 30 sec; annealing at 54°C or 58°C for the 5' or 3' end regions respectively; and 30 sec and 72°C for 8 min. Correct products yield a 9 kb or a 5.5 kb product for the 5' and 3' regions respectively.

Chimeric mice were created by injecting knockout clone 13 C57BL6 ES cells into blastocysts that were then implanted into pseudopregnant BALB/c mice. Chimeric males were identified by coat color and one male that gave almost 100% germ-line transmission of ES cell-derived material was crossed with wild-type C57BL6 females. Mice that were heterozygous for the knockout allele were identified by long-range PCR.

Multiplex genotype analysis. For routine genotype analysis DNA was extracted from the mouse tail tissue using standard purification procedures and this was screened by multiplex PCR using the following primers: one primer in the neomycin sequence, 5'-CTCTGAGCCCAGAAAGCGAAG-3' (SEQ ID NO: 14); and two primers within the NPC1L1 exon 2 sequence, NPC1L1a, 5'- GACCAGAGCCTCTTCATCAATGT-3' (SEQ ID NO: 15) and NPC1L1b, 5'-GAGAATCTGCGCTTACGAGGGA-3' (SEQ ID NO: 16) that flanked the neomycin insertion. The neomycin and NPC1L1b primer pair amplifies the knockout allele to produce a PCR product of 815 bp while the NPC1L1a and NPC1L1b primers amplify the 601 bp wildtype allele. PCR amplification used 30 cycles of denaturation at 94°C for 40 sec, annealing at 58°C for 30 sec and extension at 72°C for 1 min.

Results

Chimeric C57BL6 ES cell/BALBc mice were successfully generated and crossed with C57BL6 females. Homozygous *NPC1L1-/-* mice were identified by long-range PCR-amplification to verify that the neomycin/*NPC1L1* gene knockout cassette was correctly inserted by homologous recombination (Figure 3d). Mice were routinely screened by PCR to determine their genotype.

The resulting *NPC1L1-/-* mice were found to breed normally and showed no obvious phenotype when compared with their wild-type *NPC1L1+/+* counterparts. This was surprising considering that mice lacking NPC1 are generally sterile. These results do not exclude the possibility of subtle defects, such as those giving rise to minor abnormalities in the nervous system.

EXAMPLE 5: Analysis of Lipid Uptake and Trafficking in Wild-Type and NPC1L1 Knockout Mouse Cells

Introduction

NPC1L1 and NPC1 share a number of key structural features, including thirteen membrane spanning regions and a putative sterol sensitive motif. Accordingly, an important question is whether NPC1L1 shares some of the same functional properties as NPC1L1, specifically in the transport and movement of lipids. The present invention addresses the issue with a genetic-based approach in normal and NPC1L1-deficient mouse cells.

Methods

Generation of SV40-immortalized cell lines. Wild-type and NPC1L1 knockout mice that were 3-6 days old were euthanized in a sterile environment and liver tissue was removed and minced into 3-4 mm pieces. These were washed in PBS, transferred to 1ml of ice-cold 0.25% trypsin/100 mg tissue and incubated at 4^0C for 16 hours. The trypsin was removed and the tissue incubated at 37°C for 10-30 min. DMEM medium containing 10% FBS and 2 mM L-glutamine was added, the cells were dispersed by pipetting and then kept in culture until they began to proliferate. Cells were transfected with the pTTKneo plasmid as previously described (Smart et al., 2004). Clones of SV40-transformed cells were picked and expression of the SV40 antigen was confirmed by immunofluorescence analysis using an anti-SV40 T antigen monoclonal antibody (BD biosciences pharmingen, San Diego, CA).

Fatty Acid Uptake Assays. Fatty acid uptake was carried out essentially as described (Pohl et al., 2002), using wild-type and NPC1L1 knockout mouse cells grown to confluency. Briefly, cells grown in 6 well dishes were washed in PBS and then incubated at 37°C with 1ml of prewarmed DMEM medium containing 173 μM BSA:173 μM sodium oleate with 0.43 μM ^{3}H sodium oleate (23 Ci/mmol, Perkin

Elmer, Wellesley, MA). The assay was stopped by the addition of 2 ml ice-cold DMEM containing 200 μM phloretin and 0.5% BSA and the cells incubated on ice for 2 min. The cells were then washed six times with ice cold DMEM and lysed in 1 ml of 1M NaOH. Protein concentrations were determined using the fluorescamine assay (Bishop et al., 1978). Scintillation counting was used to measure the ^{3}H sodium-oleate in 100 μl of lysate. All samples were assayed in triplicate. A similar procedure was used to measure cholesterol uptake. ^{3}H-cholesterol was solubilized using cyclodextrin essentially as described (Sheets et al., 1999). Briefly, a mixture containing 110 μl of ^{14}C-cholesterol (52.9 mCi/mmol, Perkin Elmer), 1mg cholesterol and methyl-β-cyclodextrin solution (mβCD/Chol 8:1 mol/mol) was sonicated in a bath sonicator for 15 min prior to an overnight incubation at 37°C. Confluent cells were incubated with 1ml of DMEM containing 10μl of solubilized cholesterol at 37°C for 0-40 min.

NBD-Cholesterol and NBD-LacCer Uptake. The fluorescent sphingolipid NBDLacCer was obtained complexed to BSA (Molecular Probes) and incubated with subconfluent cultures in serum-free media for 5-10 min. The fluorescent probe was removed and fresh media containing serum was added. Cells were imaged live using a fluorescent microscope equipped with a CCD camera. NBD-cholesterol was complexed with cyclodextrin as described above for ^{3}H-cholesterol. The cholesterol/cyclodextrin complex was added to cells as described above for NBD-LacCer. Cells were processed and imaged as above.

Construction of mYFP-caveolin and fluorescent reporter vectors. To generate an mYFP-Caveolin fusion vector, caveolin-1 (GenBank accession number NM_001753) was amplified from a cDNA pool generated using human fibroblast mRNA, using the primers 5'-GCGAATTCTATGTCTGGGGGCAAATACGTAGA-3' (SEQ ID NO: 17) and 5'-GCGGATCCTTATAT TTCTTTCTGCAAGTTGATGCGGA-3' (SEQ ID NO: 18) Caveolin-1 was cloned at the 3' end of mYFP cDNA (described above) to generate the mYFP-Caveolin-1 fusion. The SRE-GFP vector was as previously described. To generate the DR4-GFP vectors the SRE element was removed from SREGFP and replaced by 3 copies of a DR4 element5 encoded by a double stranded oligonucleotide,

5'-
TTGGGGTCATTGTCGGGCATTGGGGTCATTGTCGGGCATTGGGGTCATTGTC
GGGCA-3' (SEQ ID NO: 19) To generate a construct of RFP driven by the human ABCA1 promoter the genomic sequence of the promoter was amplified (nucleotides -189 to +32) (Walter et al., 2002) and inserted into the pDsRed-Express vector (Clontech).

Results

To further characterize the role of NPC1L1 in lipid transport, mouse fibroblasts were isolated from *NPC1L1*+/+ (Wt) and *NPC1L1*-/- (L1) mice and were immortalized by expression of the SV40 large T antigen6. To characterize the response of these cells to changing lipid levels vectors were constructed in which the expression of GFP or RFP is controlled either by the ATP binding cassette transporter A1 (ABCA1) promoter, a dual DR4 element. or a dual sterol-regulatory (SRE) element. Expression of these constructs in the Wt and L1 cells indicated that the L1 cells are unable to express RFP driven by the ABCA1 promoter or DR4 element (Figure 2f). Both cell lines however, could express the SRE-driven GFP construct (Figure 2f) and responded identically to the LDL-derived sterol transport inhibitor U18666A. These results provided evidence that the L1 cells have a normal SRE response but they are unable to sense or regulate their lipid efflux response.

To evaluate the extent of this transport defect it was next determined whether the absorption and endocytosis of lipids at the plasma membrane was also altered. To assess cholesterol influx rates, radio labeled cholesterol was incubated with cells for 0-40 min. Both cell lines exhibited saturatable uptake but transport into the L1 cells was reduced by 30% (Figure 3a). Similarly, incubation with oleic acid revealed that L1 cells had a 5-10% decrease in uptake (Figure 3b). Next cells were labeled as above with a fluorescent cholesterol analog and chased for various lengths of time. Initially, cholesterol decorates the plasma membrane of both Wt and L1 cells in a punctate manner (Figure 3c). However, by 180 min, in Wt cells, NBD-cholesterol was localized at a single intracellular site, presumably Golgi, whereas in the L1 cells cholesterol accumulated in multiple intracellular pools (Figure 3c).

In addition, incubation with the fluorescent sphingolipid NBD-lactosylceramide indicated that in addition to differences in the transport of cholesterol and fatty acids, L1 cells are also defective in their transport of sphingolipids. After 15 min of chase, NBD-lactosylceramide localized to the Golgi apparatus of Wt cells and this localization was complete by 40 min (Figure 3d). However, in L1 cells NBDlactosylceramide was trapped in intracellular vesicular structures and did not reach the *Golgi* complex even after 120 min of chase (Figure 3d). Intriguingly, this phenotype has recently been described in NPC1-defective cells (Puri et al., 1999), lending further support to the notion that NPC1 and NPC1L1 may perform similar functions.

The differences in lipid endocytosis between Wt and L1 cells suggested that the lack of NPC1L1 activity causes a generalized lipid transport block that may involve deregulation of caveolae formation and/or internalization. The caveolin family of small transmembrane proteins includes caveolin-1/VIP21, caveolin-2, and a muscle-specific isoform caveolin-3. Caveolin-1 spans the plasma membrane twice forming a hairpin structure on the surface and forms homo- and hetero-oligomers with caveolin-2. Caveolins are the principle constituents of caveolae (small non-clathrin coated invaginations in plasma membrane). They preferentially associate with inactive signaling molecules such as Src and Ras family proteins and have been proposed to act as a scaffold for the assembly of signaling complexes. Caveolin-1 colocalizes and associates with the integrin receptors *in vivo*. It regulates binding of the Src family kinases to the integrin receptors to promote adhesion and anchorage-dependent growth. Other proposed functions for caveolins include regulation of cell proliferation and tumor suppression.

Expression of a mYFPcaveolin construct showed that in Wt cells caveolin localizes in a perinuclear *Golgi* area and in peri-plasma membrane ring structures (Pohl et al., 2004; Westerman et al., 1999) (Figure 3e). In striking contrast, the caveolin L1 cells appears to be trapped at the plasma membrane (Figure 3e), suggesting that lack of NPC1L1 activity causes its aberrant trafficking or mislocalization. The inability of L1 cells to endocytose caveolae may partially explain their multiple lipid transport defects.

To determine whether NPC1L1 is active in caveolae colocalization studies were carried out between mYFP-caveolin and NPC1L1-mCFP. No significant colocalization between the two proteins was detected (data not shown) suggesting that the effects seen in L1 cells are not a direct effect of the lack of NPC1L1 activity in caveolae.

5

EXAMPLE 6: Studies of Lipid Physiology in Wild-Type and NPC1L1-Knockout Mice

Methods

10 ***Animal Care.*** All mice were housed in the Mount Sinai animal care facility with controlled humidity and temperature levels and with 12 hour alternating light and dark cycles. Experiments were carried out according to protocols approved by the Institutional Animal Care and use Committee (IACUC). For colony maintenance the mice were given a regular chow diet (Lab Diet rodent diet 20, PMI Nutritional 15 International Richmond, IN) and water *ad libitum*. For studying the effects of an atherogenic diet the Paigen high cholesterol, high fat diet1 was administered (Research Diets, cat. no. D12336) and contained 12.5 gm% cholesterol, 5 gm% sodium cholic acid and a fat content of 35 kcal%. The matched low fat diet (cat. no. D12337) contained 0.3 gm% cholesterol, no cholic acid and a fat content of 10 kcal%.

20 ***Plasma lipid Assays.*** For plasma lipid assays, mice were given the high and low cholesterol diets for 14 weeks and then fasted for 16 hours. They were euthanized using a lethal dose of the anesthetic Avertin and total body blood was withdrawn from the inferior vena cava. Four male and four female mice were used for each diet.

Histology. Livers from mice fed a high cholesterol diet were excised and fixed 25 in 4% paraformaldehyde in PBS. They were embedded in paraffin, deparaffenized, rehydrated and 5□m sections were stained using 0.1% hematoxylin and 0.25% alcoholic eosin. These were mounted in Permount and examined using a Nikon light microscope.

Results

The *NPC1L1*+/+ and *NPC1L1*-/- mice were placed on a high cholesterol diet for 14 weeks. When serum lipid levels from these mice were evaluated, no significant differences were observed between *NPC1L1*+/+ and *NPC1L1*-/- mice on normal low cholesterol diet. As expected, Wt mice on the high fat diet exhibited an increase in total cholesterol and LDL-cholesterol and a decrease in their triglycerides whereas HDL-cholesterol was similar to those of animals kept on the low fat diet. However, the *NPC1L1*-/- mice given a high fat diet showed no elevation in total and LDL-cholesterol and in fact showed a significant decrease in total cholesterol. These animals had a decrease in HDL levels and had similar triglyceride levels to mice kept on the low fat diet. In addition, *NPC1L1*-/- mice on the high fat diet had a significant decrease in plasma glucose compared to *NPC1L1*+/+ mice, which has a small but significant increase in plasma glucose (assayed following overnight fasting).

Histochemical analysis of liver tissues from these animals showed that *NPC1L1*+/+ mice on the high fat diet had larger, fat-laden livers, while livers from the knockout mice were normal but smaller than the Wt high-fat livers, indicating that these animals resisted the diet-induced fatty liver. Liver sections from *NPC1L1*+/+ and *NPC1L1*-/- mice confirmed the lipid-laden status of the *NPC1L1*+/+ livers and the resistance of *NPC1L1*-/- animals to this diet induced lipid accumulation. Also, gall bladders from Wt and *NPC1L1*-/- mice on the high fat diet were dramatically different with *NPC1L1*+/+ gall bladder tissues, showing obvious signs of lipid-induced cholestasis that were absent in the *NPC1L1*-/- mouse. Together, these data show that inactivation of the NPC1L1 protein has a protective effect against diet-induced hypercholesterolemia in these animals and suggest that NPC1L1 has a critical role in regulating lipid or glucose metabolism.

EXAMPLE 7: Screening Assays for the Identification of NPC1L1 Modulators

A number of assays have been developed for the monitoring of NPC1L1 function. These assays include, for example, prokaryotic *in vivo* assays; prokaryotic *in vitro* assays; eukaryotic *in vivo* assays; and reconstitution.

All of these assays are amenable to high-throughput screening and offer four diverse ways for screening small molecule libraries. Below is a description of the various approaches.

Prokaryotic assay

NPC1L1 has been successfully expressed in a prokaryotic host (*E. coli*). In these bacteria the protein is imbedded into the inner membrane. The engineering of the expression construct involved the replacement of the NPC1L1 ER-targeting signal sequence with that of the *E. coli* protein OmpA1. An IPTG-inducible promoter drives the expression of NPC1L1.

The expression host is a derivative of *E. coli* K12. This host was engineered to lack the prokaryotic permease AcrB (a permease that has homology with NPC1L1). The host was then engineered to also lack a second component of this system a protein called TolC, by homologous recombination deletion. This host has a tremendous advantage for our studies since the AcrB/TolC system in *E. coli* is very efficient and can work to mask or confuse the results of transporter expression studies.

In vivo: Using the above host the transport of specific substrates is able to measured by looking at growth rates and/or resistance to various compounds added to the growth media since NPC1L1 transports these substances into the bacteria where they exert a toxic effect. These assays can be done on semisolid or liquid media.

In vitro: Using the above cells we can produce membrane vesicles of the inner membrane that contain the NPC1L1 protein. These vesicles can be produced with the NPC1L1 protein facing the inside (IO; inside out) or the outside (RO; right site out) of the vesicle. This is extremely useful since one can measure material going into the vesicles or coming out of the vesicles depending on need.

Thus, one can use the above system as a high throughput screening for either activators (agonists) or inhibitors of NPC1L1.

Eukaryotic *in vivo*:

Mammalian: Cell-lines have been generated that express NPC1L1 or and cell-lines have been generated that lack NPC1L1 activity. Cells lacking NPC1L1 exhibit a

number of differences with cells that express NPC1L1. These differences are measurable and can be monitored in live cells by fluorescence detection and/or microscopy. Thus, the effects or activity of various small molecules on the activity of NPC1L1 can be evaluated in a high-throughput screening system.

Baculovirus: A very high-level expression system has been produced based on baculovirus that expresses NPC1L1 tagged at the C-terminus with a dual histidine-HA tag in insect cells. This provides an efficient and quick way to purify large quantities of recombinant NPC1L1 for reconstitution studies/screening (see below). In addition, these cells can be used to confirm results or candidate molecule identified by one of the methods described above.

Reconstitution:

Purified NPC1L1 from insect cells: Purified material from the above (baculo) can be used to form vesicles *in vitro* using various lipid compositions including the one that NPC1L1 resides in (Golgi membranes). Fluorescent or radioactive probes can be incorporated into the membrane of these vesicles or captured into their interior hydrophilic core. Probes will be identified on their ability to change location within these vesicles dependent on the activity of NPC1L1. And therefore, their movement can be monitored in the presence of compounds that change (increase or decrease) the activity of NPC1L1.

A mammalian cell assay for screening potential NPC1L1 is described herein (see ricin assay as described in Example 10, below) and a prokaryotic system for screening potential NPC1L1 inhibitors is described in Example 8.

EXAMPLE 8: <u>Assay for Inhibitor Screening for NPC1 and NPC1L1 and Identification of 4-phenylpiperidines as potent inhibitors of NPC1</u>

In order to devise an assay for inhibitor screening a system where some potential activity of NPC1 or NPC1L1 can be detected and monitored is needed. Also, further complications are added by the fact that expression of these proteins in mammalian cells is usually not tolerated and sometimes lethal.

Methods

The present inventors have devised a prokaryotic expression system for both NPC1 and NPClL1 based on the expression of these proteins with prokaryotic secretion signals for targeting the *E. coli* inner membrane. The engineering of the expression construct involved the replacement of the NPC1 and NPC1L1 ER-targeting signal sequences with that of the *E. coli* protein OmpA1. An IPTG-inducible promoter drives the expression of NPC1 and NPC1L1. This system for expression of NPC1 has been described by the inventors (see Davies, Chen and Ioannou, *Science* 290: 2295-98, 2000).

In addition, hosts have been engineered to allow for the efficient detection of any potential activities as described in Example 9, below.

The expression host is a derivative of *E. coli* K12. This host was engineered to lack the prokaryotic permease AcrB (a permease that NPC1 and NPC1L1 have homology with), and was a gift from Dr. Tomofusa Tsuchiya (*Antimicrob. Agents and Chemoth.* 42: 1778, 1998). The host was engineered to lack a second component of this system a protein called TolC, by homologous recombination deletion (Figure 5). This host has a tremendous advantage for these studies since the AcrB/TolC system in *E. coli* is a very efficient drug efflux system and can work to mask or confuse the results of transporter expression studies.

The final improvement made was introducing into these strains mutations that make their outer membrane leaky. The *E. coli* outer membrane is a strong barrier of lipophilic molecules and thus prevents any assays to be carried out that involve lipophilic substrates. Since the predicted substrates of NPC1 and NPC1L1 are lipophilic it is critical to engineer a strain that has a leaky outer membrane. In this manner lipophilic molecules can cross the outer membrane so that they can interact with the expressed NPC1 and NPC1L1 proteins residing on the inner membrane of the bacteria.

Utilizing this bacterial host it was discovered that these mutants are unable to grow in the presence of 5 mM concentration of a short chain fatty acid (decanoate; a 10 carbon length fatty acid). However, bacteria expressing NPC1 are able to overcome this block and grow in the presence of decanoic acid. In one type of assay bacteria are

plated onto a dish to form a lawn. Small filter disks (about 8mm diameter) are soaked in decanoate and placed onto the bacterial lawn. Dishes are incubated overnight at 37°C and inspected the next morning. The substance (decanoate or other test material) diffuses from the filter in a radial manner into the bacterial lawn and will inhibit bacterial growth. The diameter of the inhibition ring (around the filter) will be directly related to the sensitivity of the bacteria to the test substance; the more resistant the bacteria are to the test substance the closer to the filter they will grow forming a smaller diameter ring.

This assay works equally well in liquid cultures; decanoate is added to liquid cultures and bacteria are grown at 37 °C with shaking for 4-6 hours. At the end of the incubation period an optical density measurement at 600 nm (OD_{600}) determines the ability of the culture to grow. Using the above cultures it was determined that control bacteria grew at an OD=0.9 whereas NPC1-expressing bacteria grew to saturation of OD>3.0.

Results

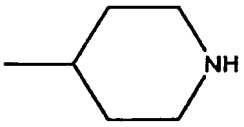
The above assays were used to search for inhibitors of NPC1 and NPC1L1. On the plate assay various inhibitors, as set forth below, were added to the cultures before plating and searched for molecules that did not interfere with the growth of control bacteria in the presence of decanoate. In the NPC1-expressing cells an increase in the diameter of the growth inhibition ring was observed, suggesting that the NPC1 protein is inhibited and leads to these bacteria regaining their sensitivity to decanoate. A number of molecules were screened and a number of candidate inhibitors identified (set forth below). The two most promising candidates were validated in mammalian cell cultures.

Cells were treated with these inhibitors and cholesterol storage was monitored. Cells treated with molecule #5 (4-butyryl-4-phenylpiperidine hydrochloride-see) overnight. In the presence of these inhibitors, mammalian cells should exhibit a disease phenotype (the human lipidosis Niemann-Pick C is due to a deficiency of NPC1). Cells from NPC1 patients store cholesterol in their lysosomes, which can be easily visualized by staining cells with a fluorescent probe that recognizes cholesterol.

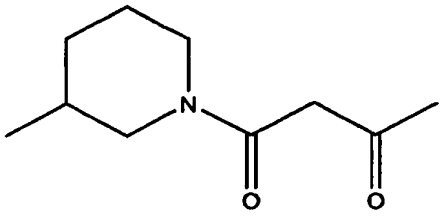
Results are shown in Figure 6 and Figure 7. No significant staining for lysosomal cholesterol can be seen in normal human fibroblasts (Figure 6A). However, the same fibroblasts incubated overnight with inhibitor #5 have distinguishable lysosomes filled with cholesterol (Figure 6B).

Molecule #2 (4-methylpiperidine) was a weaker NPC1 inhibitor, although fibroblasts treated with this inhibitor still exhibit cholesterol-filled lysosomes (Figure 7A). Molecule #1 (4-phenyl-4-phenylpiperidine hydrochloride) did not demonstrate any NPC1 inhibition, as shown by an absence of cholesterol build-up in the lysosomes (Figure 7B). The molecules identified as potential NPC1 inhibitors may also be effective as NPC1L1 inhibitors. For example, Molecule #1 (4-phenyl-4-phenylpiperidine hydrochloride), has been identified as an inhibitor of NPC1L1, even though it did not demonstrate any NPC1 inhibition.

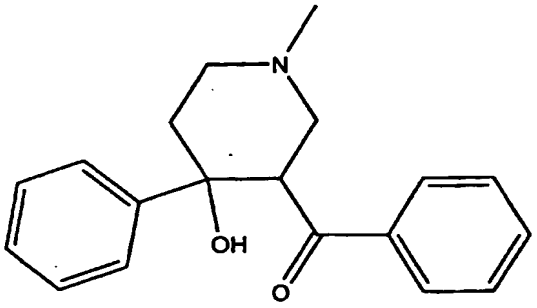
Candidate Inhibitors Identified Using the Above-Described Assay:



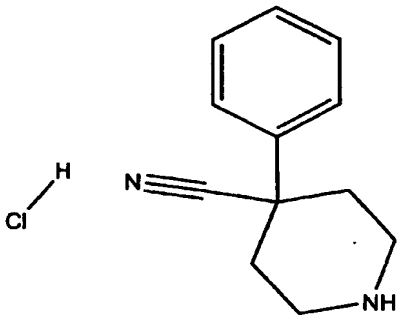
4 methylpiperidine #2



1-ACETOACETYL-3-METHYLPIPERIDINE

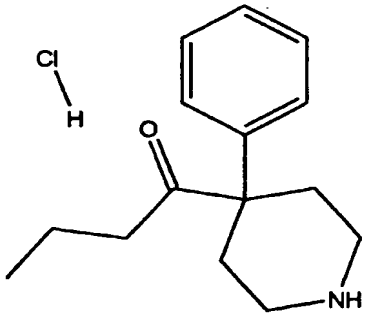


3-BENZOYL-4-HYDROXY-1-METHYL-4-PHENYLPIPERIDINE #3

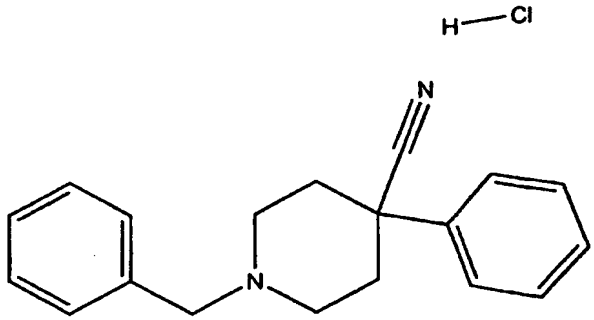


4-Phenyl-4-piperidinecarbonitrile Hydrochloride #1

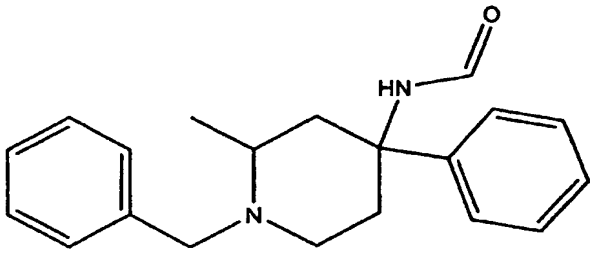
5



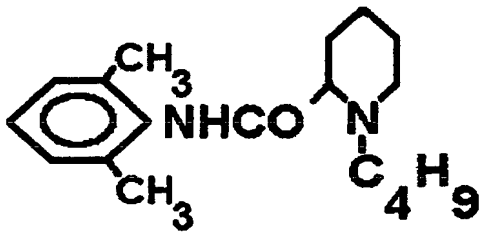
4-BUTYRYL-4-PHENYLPIPERIDINE HYDROCHLORIDE #5



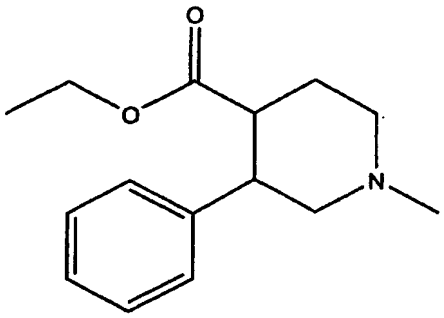
1-BENZYL-4-CYANO-4-PHENYLPIPERIDINE HYDROCHLORIDE #4



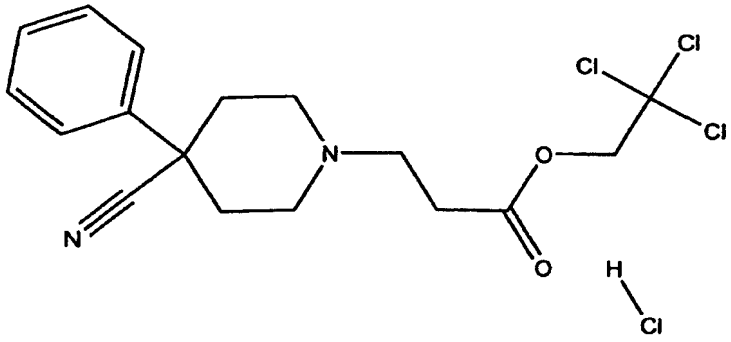
1-BENZYL-4-FORMAMIDOMETHYL-4-PHENYLPIPERIDINE #6



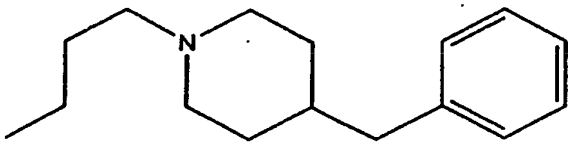
Bupivacaine hydrochloride B5274



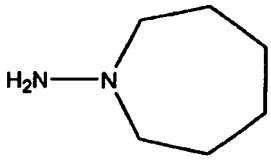
ETHYL 1-METHYL-3-PHENYL-4-PIPERIDINECARBOXYLATE



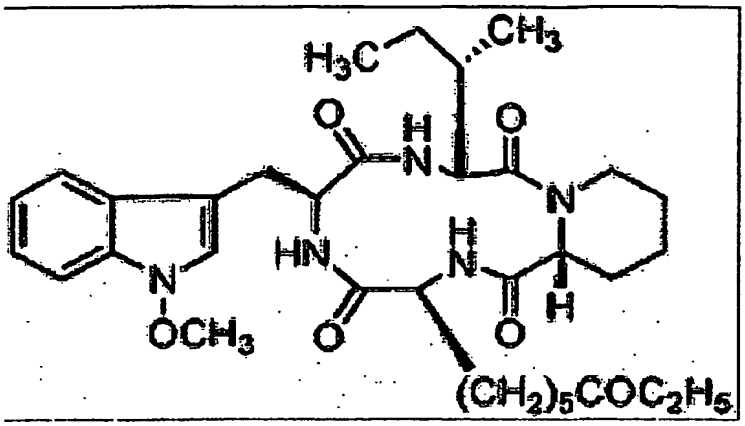
2,2,2-TRICHLOROETHYL 4-CYANO-4-PHENYL-1-PIPERIDINEPROPIONATE HCL



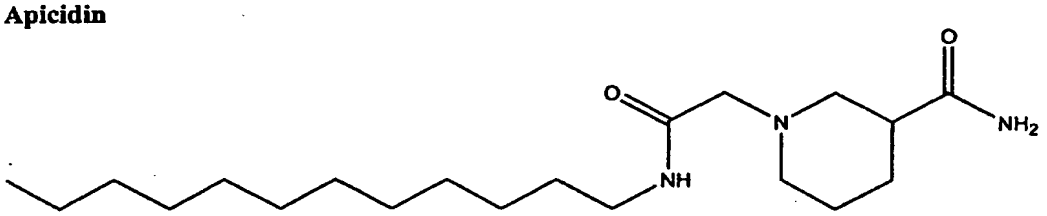
4-BENZYL-1-BUTYLPIPERIDINE



1-Aminohomopiperidine



Apicidin



3-CARBAMOYL-N-DODECYL-1-PIPERIDINEACETAMIDE

EXAMPLE 9: <u>Engineered *E. coli* Hosts for High-Level Expression of Mammalian Transporters.</u>

Expression of NPC1 in bacteria as previously described by the inventors (Davies et al., 2000 *Science* 290, 2295-2298) was limited by the fact that *E. coli* bacteria have a number of efflux pumps that belong in the Resistance-Nodulation-Division (RND) family. These pumps transport molecules away from the *E. coli* cytosol in direct opposition to the direction of transport by NPC1 and NPC1L1. This in turn complicates analysis of experimental data generated in this system. Thus, an AcrB mutant strain has been obtained which lacks one of the major RND permeases part of the AcrA, AcrB and TolC complex.

First, using this strain the TolC gene has been mutated by homologous recombination using the approach recently described (Link et al., 1997 *J. Bacteriology*

179, 6228-6237). The TolC gene forms the channel on the *E. coli* outer membrane and it is shared by most of the RND permeases in *E. coli*. Thus, inactivating this gene effectively inactivates most if not all, *E. coli* RND permeases.

Second, following construction of the double AcrB, TolC mutant strain, these bacteria were mutagenized and selected for strains with a "leaky" outer membrane similar to the previously described selection procedure (Davies et al., 2000 *Science* 290, 2295-2298). This mutagenesis produced an AcrB/TolC/permeable strain.

Third, this triple mutant, (AcrB/TolC/Perm), was used to select for expression of large transmembrane proteins. This selection is accomplished by allowing NPC1-expressing and NPC1L1-expressing bacteria to spontaneously mutate on agar plates (as described by Miroux and Walker, 1996 J Mol Biol *260*, 289-298; Shaw and Miroux, (2003). A general approach to heterologous membrane protein expression in escherichia coli. In Membrane Protein Protocols, B. S. Selinsky, ed. (Totowa, NJ, Humana Press), pp. 23-35). Colonies that can grow and continuously express NPC1 and NPC1L1 were isolated and cured of the NPC1 or NPC1L1 expression plasmids. This selection produced two strains:

a. AcrB/Tolc/Perm/N1; and

b. AcrB/Tolc/Perm/L1.

EXAMPLE 10: <u>NPC1L1 Assay Based on Ricin Endocytosis</u>

Following the observation that human liver has the highest expression of NPC1L1, the human liver derived cell line Huh7 was characterized. These cells express significant amounts of NPC1L1 as seen by mRNA and protein levels and were chosen for subsequent studies.

First, stable clones were generated that expressed higher levels of NPC1L1 by introducing the human NPC1L1 cDNA into these cells. About 30 clones were characterized and clone number 3 had about a five-fold increase in NPC1L1 protein expression.

Next, a number of siRNAs were designed that targeted the NPC1L1 mRNA at various positions. These siRNAs were tested and it was found that two siRNAs

targeted NPC1L1 very efficiently. The sequence of these siRNAs are set forth as follows:

1165: TGGTCTTTACAGAACTCACTA (SEQ ID NO: 23)

1484: TCCGGACAATACCAGTCTCTA (SEQ ID NO: 24).

The numbers 1165 and 1484 refer to the nucleotide position of the human NPC1L1 cDNA (set forth as SEQ ID NO:21), which is the first nucleotide of each siRNA.

Below are the actual construct sequences that were included in the siRNA expression vector (commercially available from GenScript™). The sequences were cloned into a BamHI-HindIII sites.

NPC1L1 Si RNA 1165

GGATCCCGTAGTGAGTTCTGTAAAGACCATTGATATCCGTGGTCTTT

BamHI Antisense Loop Sense

ACAGAACTCACTATTTTTTCCAAAAGCTT (SEQ ID NO: 25).

Terminator

NPC1L1 Si RNA 1484

GGATCCCGTAGAGACTGGTATTGTCCGGATTGATATCCGTCCGG

BamHI Antisense Loop Sense

ACAATACCAGTCTCTATTTTTTCCAAAAGCTT (SEQ ID NO: 26).

Terminator

Both of these siRNAs were introduced into a vector and stable cell-lines were generated. More than 50 of these cell lines were characterized and four were chosen to be characterized further. Si6 was found to be the best cell-line. Si6 has greater than a 90% decrease in the NPC1L1 mRNA making this clone effectively null for NPC1L1 protein expression.

To further characterize these clones, a number of experiments were carried out using lipid uptake and various toxins to probe their transport. Fluorescent lipids ceramide, cholesterol and LacCer were incubated with cells for 60 minutes at 4°C and then chased at 37°C for 30 minutes. All lipids exhibited altered uptake and localization when compared between the NPC1L1 positive clone number 3 and the NPC1L1 negative si6 clone. In particular, there was pronounced Golgi localization of all lipids in the NPC1L1 negative si6 cells.

The endocytosis of a number of toxins such as Ricin, Diphtheria toxin and Verotoxin were then tested. In the case of ricin, the si6 cells appear to target this toxin to the Golgi much more rapidly than either the wild type cells or the clone number 3 cells. To confirm that these results are not due to something unique to clone si6, the ricin uptake experiment was repeated with other, independent siRNA clones. All of these clones, with the exception of clone siS6, which was probably not a good siRNA clone, gave the same result with respect to ricin endocytosis.

A time course experiment was then carried out to determine the optimal time for detecting these differences in endocytosis. It was determined that as early as 15 minutes following addition of the toxin, the different in endocytosis is apparent. Si6 cells show a dramatic Golgi staining with the toxin whereas the wild type and number 3 clone cells exhibit only a punctate type of staining.

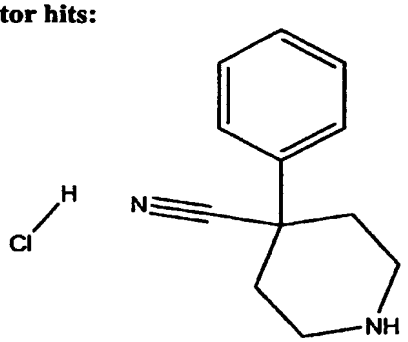
Finally, to capitalize on these differences the viability of these cells to Ricin, Diphtheria toxin and Verotoxin intoxication was tested. As predicted from the above results the si6 cells are much more sensitive to the toxins since they appear to target these toxins to their Golgi more efficiently. Si6 cells exhibit higher sensitivity to Ricin following incubation with Ricin overnight.

Alternatively, higher amounts of toxin (5 ug/ml) were incubated with the cells for different amounts of time. With this approach, similar to the above, a two-fold difference in Ricin sensitivity was seen.

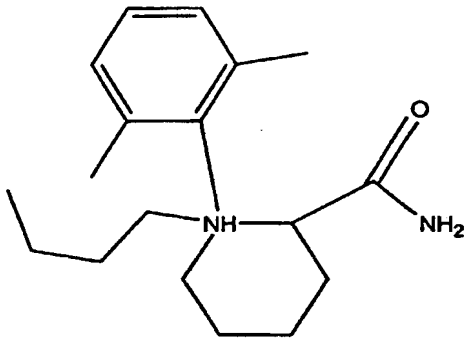
In conclusion, the number 3 clone and Ricin intoxication can be used in an assay to measure an increase in the number 3 clone's sensitivity to Ricin based on NPC1L1 inhibition.

The above described mammalian cell assay has been used to screen a library of 3,000 compounds. Molecules that are inhibitors of NPC1L1 activity have been identified (see inhibitors below). A prokaryotic system for screening potential NPC1L1 inhibitors is also described herein (see Example 8).

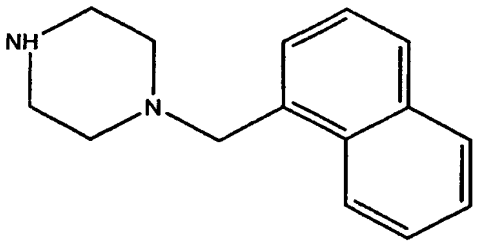
Inhibitor hits:



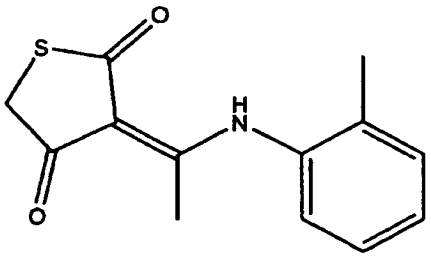
4-Phenyl-4-piperidinecarbonitrile Hydrochloride; and



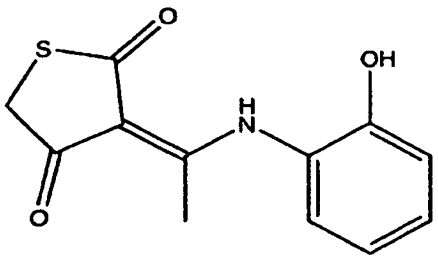
1-Butyl-N-(2,6-dimethylphenyl)-2 piperidinecarboxamide; and



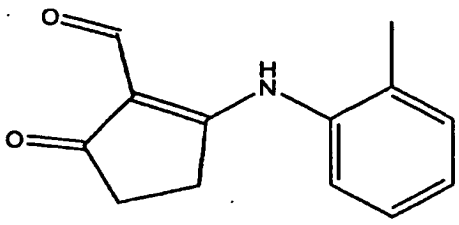
1-(1-Naphthylmethyl)piperazine



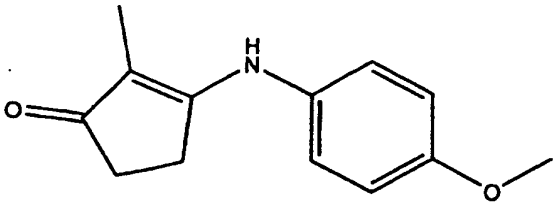
3-{1-[(2-methylphenyl)amino]ethylidene}-2,4(3H,5H)-thiophenedione; and



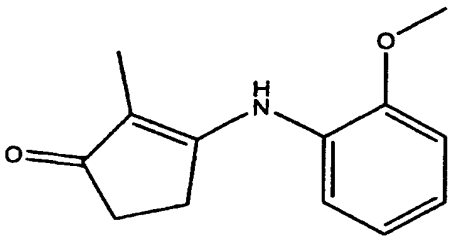
3-{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H,5H)-thiophenedione; and



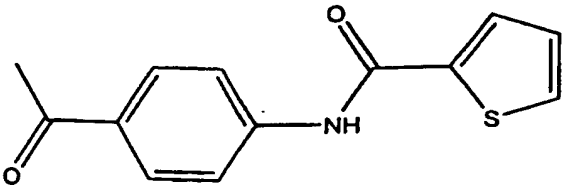
2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one; and



3-[(4-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one; and



3-[(2-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one; and



N-(4-acetylphenyl)-2-thiophenecarboxamide .

EXAMPLE 11: Assay of NPC1L1 Function by Measuring Expression of the NPC1L1 Promoter.

The inventors observed that the NPC1L1 knockout mice described herein have high levels of truncated NPC1L1 mRNA. This suggests that lack of NPC1L1 activity induces expression of NPC1L1. This observation can therefore be used to develop an assay for screening for NPC1L1 inhibitors.

Reporter vectors were constructed that place expression of the luciferase gene under the control of the human NPC1L1 promoter or the mouse NPC1L1 promoter. To validate this, the human construct was transfected into three human liver cell lines.

The promoter sequences of human and mouse NPC1L1 are set forth as SEQ ID NO: 27 (human) and SEQ ID NO: 28 (mouse). These sequences are in the constructs driving the expression of luciferase in vector pGL3 (Promega Corp™). These sequences also include the start codon and a short piece of protein coding region from the 5' end of the genes and are cloned in-frame with firefly luciferase, thus creating luciferase with a short piece of NPC1L1 fused to it's 5' end. The start codon region is included because a potential transcription factor, YY1, is known to be involved in the regulation of several key lipid homeostasis genes; in the human NPC1L1 promoter the transcription factor site covers the ATG in an antisense orientation and may possibly inhibit transcription of the gene from this start site.

As predicted, expression of luciferase in Wt Huh7 (wild type; human liver) cells was detectable since these cells express NPC1L1 and therefore are expected to also express luciferase driven by the NPC1L1 promoter. When the construct was introduced into the Huh7 cells where expression of NPC1L1 is inhibited by an siRNA (Si6 as described above), expression is up regulated. In contrast, expression in cells that overexpress NPC1L1 (L1 3+) is down regulated compared to wild type cells (Wt) and even more so compared to the cells that do not express NPC1L1 (Si6 cells).

These results indicate that NPC1L1 is unique in that it regulates its own expression. That is, when cells sense that there is lack of NPC1L1 activity the cells up-regulate the NPC1L1 promoter and when levels of NPC1L1 protein rise the cells down-regulate NPC1L1 expression. Thus, the L1 3+ cell-line can also be used for

screening NPC1L1 inhibitors. Inhibitors of NPC1L1 induce expression of the luciferase gene driven by the NPC1L1 promoter to the levels detected in the Si6 cells, *e.g.*, about 4-5 fold higher.

The inhibitors identified using the ricin intoxication assay (Example 10) were tested in utilizing the above assay whereby upregulation of the NPC1L1 promoter was used to detect the inhibition of the NPC1L1 protein. As shown in Figure 8, 4-Phenyl-4-piperidinecarbonitrile Hydrochloride (#1), (1-Butyl-N(2,6-diemethylphenyl)2 piperidine carboxamide) #7, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione and gave a positive signal compared to control (none). Note that Ezetamibe did not inhibit NPC1L1 in this assay.

EXAMPLE 12: Comparison of NPC1L1 (-/-) Knockout and C57BL6 Wild-Type Mice Fed a High Fat Diet

Wild-type C57BL6 mice are known to be susceptible to diet induced obesity, followed by the development of type II (non-insulin dependent) diabetes. Administration of a diabetogenic high fat diet can induce these symptoms in wild-type C57BL6 mice.

Obesity is strongly associated with diabetes and as the mice become progressively more obese there is an increase in lipid deposition in adipose tissue, along with ectopic deposition of lipid in key peripheral tissues such as skeletal muscle, the liver and pancreas. Elevated amounts of plasma lipids, such as fatty acids are also observed. The peripheral tissues eventually fail to respond to insulin, leading to insulin resistance, glucose intolerance and elevated plasma glucose. The pancreatic β-cells attempt to compensate for the insulin resistance and glucose intolerance by producing more insulin, leading to hyperinsulinemia. Overt diabetes occurs when the pancreatic β-cells fail to secrete adequate amounts of insulin to lower plasma glucose levels and pancreatic cell damage occurs.

Under normal conditions, insulin regulates glucose by stimulating glucose uptake and metabolism in adipose and skeletal muscle tissues. It also inhibits gluconeogenesis in the liver. In the pre-diabetic and in patients with overt diabetes,

this regulation is impaired so that plasma glucose can no longer be effectively maintained at the required levels.

The studies below compare the effect of the NPC1L1 gene knockout (-/-) with wild-type C57BL6 mice that become obese and develop type II diabetes, during administration of a high fat diet. Mice that were 7-8 weeks of age were placed on a high fat diet for these studies.

The NPC1L1 (-/-) knockout mice were protected against the diet-induced obesity and diabetic symptoms observed in wild-type (wt) C57BL6 mice. Therefore, inhibitors of NPC1L1 may be useful for the treatment and/or prevention of obesity and diabetes.

1. **Body weight of two sets of mice fed a high fat diet**

The following experiments show that whilst the wild-type (wt) C57BL6 mice become obese when fed a high fat diet, the NPC1L1(-/-) knockout mice resist the development of obesity. Data is from two independently analyzed sets of mice identified as mouse set 1 and mouse set 2.

In the first experiment, NPC1L1 gene knockout (-/-) and wild-type (wt) mice were fed a high fat diet for 0-245 days and weighed on a weekly basis for most of the time-course. There were 5 knockout mice and 6-7 wild-type mice used in this experiment.

As shown in Figure 9A, the wild-type mice became obese whilst the knockout mice resisted the weight gain. By 245 days the knockout mice had an average weight of 32.5g whilst the wild-type mice were 55.4g.

In a second experiment, NPC1L1 gene knockout (-/-) and wild-type (wt) mice were fed a high fat diet for 0-95 days and weighed on a weekly basis for most of the time-course. There were 7 knockout mice and 7 wild-type mice used in this experiment.

As shown in Figure 9B, the wild-type mice became obese whilst the knockout mice resisted the weight gain. By 245 days the knockout mice had an average weight of 25.3g whilst the wild-type mice were 45.4g.

2. **Glucose Tolerance tests on mice**

The data below shows that on a regular chow diet, at 7 weeks of age, the wild-type (wt) C57BL6 and NPC1L1(-/-) knockout mice have a normal and similar ability to clear blood glucose.

When fed the high fat diet, the NPC1L1(-/-) knockout mice, although showing slightly impaired glucose tolerance, are able to effectively regulate their blood glucose, in contrast to the wild-type mice, which show classic glucose intolerance at both 102 and 262 days of high fat diet administration.

After weaning, 7 wild-type and 5 knockout mice (age-matched) were fed a regular chow diet. At 7 weeks of age the mice were fasted overnight and then injected intraperitoneally with glucose. Blood glucose was measured from 0-120 min. There is no significant difference in the glucose tolerance of these wild-type and NPC1L1 (-/-) knockout mice as both show efficient clearance of excess blood glucose (see Figure 10).

In a second experiment, mice were placed on a high fat diet at 7-8 weeks of age and, after 102 days of feeding the high fat diet, glucose tolerance was tested in 6 wild-type and 5 gene knockout mice. The mice were fasted overnight and then injected intraperitoneally with glucose. Blood glucose was measured at 0-240 min after injection. The wild-type mice are significantly intolerant to intraperitoneal glucose injection, with slow clearance. In contrast, the gene knockout mice effectively clear the injected glucose. The glucose intolerance observed in the wild-type mice is a sign of the onset of type II diabetes and is likely to be associated with the weight gain seen in these mice. The gene knockout mice seem to be protected against this symptom of diabetes (see Figure 11A).

In a third experiment, mice were placed on a high fat diet at 7-8 weeks of age and, after 262 days of feeding the high fat diet, glucose tolerance was tested in 6 wild-type and 5 gene knockout mice. The mice were fasted overnight and then injected intraperitoneally with glucose. Blood glucose was measured at 0-240 min after injection. At 262 days of feeding on a high fat diet the wild-type mice were significantly more intolerant to intraperitoneal glucose injection, with severely slowed

clearance, compared with the NPC1L1 (-/-) gene knockout mice, which effectively reduce the elevated glucose. The glucose intolerance observed in the wild-type mice is indicative of type II diabetes. The NPC1L1 (-/-) gene knockout mice, although not completely normal in their glucose clearance time, are not nearly as severely affected as the wild-type mice (see Figure 11B).

3. **Insulin Tolerance test in mice**

Normally, insulin stimulates glucose uptake and metabolism in adipose and skeletal muscle tissues as well as inhibiting gluconeogenesis in the liver, thus lowering blood glucose levels. The data below shows that when insulin is administered to the wild-type C57BL6 mice fed a high fat diet there is little effect on the blood glucose levels in these mice, indicating that they have become intolerant to the effects of insulin in lowering blood glucose. The NPC1L1(-/-) knockout mice respond to the insulin administration with a decrease in blood glucose, as expected in insulin responsive animals.

In a first experiment, mice were fed a high fat diet for 105 days (7 wild-type and 7 knockout mice). After a 3 hour fast, mice were injected intraperitoneally with insulin and their blood glucose was measured. The decrease in blood glucose caused by insulin administration was clear in the NPC1L1 (-/-) gene knockout mice, with a rapid decrease in glucose levels. In the wild-type mice there was a muted, almost non-existent response to insulin injection as the glucose levels remained high (see Figure 12A). This insulin resistance observed in the wild-type C57BL6 mice is characteristic of mice in a pre-diabetic or overtly diabetic state.

In a second experiment, mice were fed a high fat diet for 252 days (6 wild-type and 5 knockout mice). After a 3 hour fast, mice were injected intraperitoneally with insulin and their blood glucose was measured. As at 105 days, the decrease in blood glucose caused by insulin administration was clear in the NPC1L1 (-/-) gene knockout mice, with a decrease in glucose levels. In the wild-type mice there was a muted, almost non-existent response to insulin injection as the glucose levels remained high (see Figure 12B). This insulin resistance observed in the wild-type C57BL6 mice is characteristic of mice in a pre-diabetic or overtly diabetic state.

4. **Insulin measurements in mice injected with glucose**

In a first experiment, glucose was injected intraperitoneally into 7 wild-type and 7 NPC1L1 (-/-) gene knockout mice that had been fed a high fat diet for 72 days and then fasted overnight. Plasma insulin was measured at 0-30 min. In the knockout mice the pre-injection plasma insulin was low and the increase in insulin caused by glucose injection was presumably short-lived as it was not detected at 15 minutes, the first measurement post-glucose injection, results that would be expected in non-diabetic mice (see Figure 13A). The wild-type mice have hyperinsulinemia and the elevated insulin levels are maintained throughout the course of the experiment and this is characteristic of a pre-diabetic and diabetic disease state.

In a second experiment, glucose was injected intraperitoneally into 6 wild-type and 5 NPC1L1 (-/-) gene knockout mice that had been fed a high fat diet for 220 days and then fasted overnight. Plasma insulin was measured at 0-30 min. As at 72 days, in the knockout mice the pre-injection plasma insulin was low and the increase in insulin caused by glucose injection was presumably short-lived as it was not detected at 15 minutes, the first measurement post-glucose injection, results that would be expected in non-diabetic mice (see Figure 13B). The wild-type mice have hyperinsulinemia and the elevated insulin levels are maintained throughout the course of the experiment and this is characteristic of a pre-diabetic and diabetic disease state.

5. **Plasma lipoprotein profiles in the mice at 120 and 268 days of high fat diet**

Plasma lipid profiles were analyzed in wild type and NPC1L1(-/-) mice. The knockout mice significantly lower plasma LDL and HDL and total cholesterol than the wild-type mice. The plasma triglyceride levels were similar in both groups (see Figures 14A and 14B).

EXAMPLE 13: Comparison of Food Intake of NPC1L1 (-/-) Knockout and C57BL6 Wild-Type Mice

Food intake of mice lacking NPC1L1 (NPC1L1 knockout mice) has been investigated by the inventors. It has been found that there is no difference between

wild-type and knockout mice with respect to the amount of food consumed. This indicates that lack of NPC1L1 (or inhibition of NPC1L1) does not suppress appetite.

Since NPC1L1 appears to regulate the flow of lipids (and possibly other nutrients) from the plasma membrane (uptake) to the various cellular organelles such as Golgi and ER it was hypothesized that lack (or decreased) NPC1L1 activity could have a number of effects on cellular homeostasis: 1) limit the amount of nutrients (lipids, proteins, sugars) that become available for cellular processes, 2) alter signaling cascades that tell the cell to behave as if nutrients are plentiful, and 3) stimulate a limited nutrient response.

However, when mice are challenged with a high fat diet (60 kcal% fat; Diet D12492, available from Research Diets, Inc.™, New Brunswick, NJ) the results are interesting. In the beginning stages of the high fat diet, the NPC1L1 knockout mice are eating less (about 60% of the wild-type mice). As they are challenged longer >90 days their intake becomes similar to wild-type mice. Importantly, even after 90 days, the knockout mice still do not gain as much weight as the wild-type animals (see Figure 18).

EXAMPLE 14: White Adipose Tissue Has Significant Expression Levels of NPC1L1

Previous real-time PCR data have shown that NPC1L1 is elevated in the small intestine of both mice and humans and in addition, is high in the human liver. The data described herein shows that adipose tissue expresses a significant amount of NPC1L1. Since the absence of NPC1L1 is protective against obesity and type II diabetes and adipose tissue plays a role in the development of both of these diseases, finding significant expression in these tissues is of considerable interest.

NPC1L1 transcript was measured by semi-quantitative real-time PCR, normalized to β-actin expression. As shown in Figure 15, in mouse white adipose (gonadal) tissue, NPC1L1 is expressed at 9% of the amount detected in the small intestine, which has the most abundant expression of NPC1L1. This is a significant amount compared with other tissues (for example, pancreas has only 2% of small the

amount found in the small intestine). The pre-adipocyte mouse cell line 3T3L1 does not express NPC1L1.

NPC1L1 transcript was measured by semi-quantitative real-time PCR in mouse white (gonadal) adipose (WAT) and interscapular brown adipose tissue (IBAT), normalized to β-actin expression. As shown in Figure 16, expression of NPC1L1 is higher in white adipose tissue and the amount in brown adipose is 42% of that found in the white tissue.

NPC1L1 transcript was also measured by semi-quantitative real-time PCR in human liver and white adipose tissue, normalized to β-actin expression. As shown in Figure 17, the expression in human white adipose tissue was 3% of that detected in human liver. Previously, it was found that human jejunum (the highest expressing human intestine tissue) had 4% of the NPC1L1 transcript found in human liver and so a value of 3% for adipose is a significant amount of NPC1L1. Many other tissues have less than 1% of the NPC1L1 detected in liver.

EXAMPLE 15: Creation of NPC1L1 Transgenic Mice that Overexpress NPC1L1

Rationale:

The NPC1L1 knockout mouse was instrumental in deciphering the lipid transport function of this protein and its critical role in intestinal cholesterol and other lipid transport. A powerful tool in drug discovery and drug testing (to determine is a drug acts directly on NPC1L1) is a mouse that overexpresses NPC1L1. There are a number of considerations in developing such as model. First, these mice must be able to tolerate higher expression of NPC1L1 so that its expression does not cause lethality. Second, given that the mouse NPC1L1 gene is not expressed in all mouse tissues, a system must be designed that expresses the protein at high levels but only in the appropriate tissues.

The first consideration can only be determined once the transgenic mice are generated and evaluated to see if they can pass the NPC1L1 genes to their progeny. To address the second consideration the mouse complete gene (genomic sequence as

described below) was used. In this manner, the promoter and all regulatory elements are maintained and provided the tissue specificity required.

Results:

The entire mouse gene sequence of NPC1L1 was cleaved from a Bac vector, clone RP23-64P22 (from female mouse library), obtained from BacPac Resources™, Oakland CA, which contains the unordered genomic fragments given in GenBank Accession number AC079435. The complete, ordered, gene sequence is given in GenBank sequence, accession number AL607152. According to this ordered sequence (GenBank Accession number AL607152) the gene spans nucleotides 37338 (5' end) to 18610 (3' end) in an antisense orientation.

A region spanning the complete gene was excised using the restriction endonuclease enzyme MfeI, which cleaves the region from nucleotides 6656-46736, of GenBank Accession number AL607152, containing the entire NPC1L1 gene and almost 10kb of sequence upstream of the start codon and therefore including the entire NPC1L1 promoter region for regulated gene expression.

The MfeI fragment was cloned into the 6.8kb vector pSMARTVC (Lucigen Corporation™) at its EcoRI site.

The NPC1L1/pSMARTVC vector was cleaved using AscI and PmeI and a linearized NPC1L1 fragment, with short, flanking vector arms was isolated by sucrose gradient separation to allow removal of most of the pSMART vector.

The isolated NPC1L1 gene fragment was then injected into fertilized mouse eggs and these placed into pseudopregnant C57BL6 mice (Taconic™). Transgenic mice were created by incorporation of the transgene into these mice. The mice were screened by PCR amplification of both their 5' and 3' ends, using one primer that contained the NPC1L1 gene sequence and a second primer that contained the short flanking pSMART vector arm sequence.

The primers used to amplify the 5' end of transgenic NPC1L1 have the following sequence: pSMART 5' CTATACGAAGTTATGTCAAGCGG (SEQ ID NO:

30) and mNPC1L1 BAC 46043(+) CTTGCACCTGACTTCCTCATATAAG (SEQ ID NO: 31).

The primers used to amplify the 3' end of transgenic NPC1L1 have the following sequence: pSMART 3'AAAGAAGGAAAGCGGCCGCCAGG (SEQ ID NO: 32); and mNPC1L1 BAC 7568 (-) AGGAACCGTACTGAGCGCATACCAA (SEQ ID NO: 33). Therefore, presence of the 5' and 3' ends of the NPC1L1 transgene in the progeny mice was confirmed, indicating that at least one additional copy of the mouse NPC1L1 gene had been inserted.

Two transgenic mouse lines have been created and one has successfully transmitted the transgene to its offspring (3 out of 7). Both of the parental original transgenic mice have an increased body weight, compared to the average weight of C57BL6 mice (Both transgenic mice were overweight). Male mouse #2 (which has successfully produced offspring) was 34 grams at 5.5 months of age. Female mouse #6 was 37 grams at 4 months of age (no offspring)). The average weight of a normal mouse at 4-6 months of age is about 25 grams.

Also, when genotyping these mice, the DNA was prepared by proteinase K digestion to produce crude, unpurified DNA for PCR-analysis. Unusually, there appeared to be lipid floating on the top of the extract and the OD abnormal, most likely due to excess tissue lipids.

Conclusion

The NPC1L1 gene was identified, based on its structural homology to NPC1. Cell-based studies of the NPC1L1 indicate that NPC1L1 has a predominant intracellular localization, with concentration in the Golgi and ER compartments. mRNA expression profiling of NPC1L1 reveals significant differences in RNA transcript levels between mouse and man, with highest expression levels found in human liver. Isolation of the mouse NPC1L1 gene allowed implementation of a knockout model of NPC1L. Mice lacking a functional NPC1L1 have multiple lipid transport defects. Surprisingly, lack of NPC1L1 exerts a protective effect against diet-induced hyercholesterolemia. When compared with wild-type controls, NPC1L1-

deficient mice also show a different response in levels of glucose, LDL-cholesterol, and HDL-cholesterol following a shift from a low-fat to high-fat diet. Further characterization of cell lines generated from wild-type and knockout mice reveals that, in contrast to wild-type cells, NPC1L1-deficient cells show aberrations in both plasma membrane uptake and subsequent transport of a variety of lipids, including cholesterol, fatty acids, and sphingolipids. Furthermore, cells lacking NPC1L1 reveal aberrant caveolin transport and localization, suggesting that the observed lipid defects may result from an inability of NPC1L1 to properly target and regulate caveloin expression. Furthermore, comparison of NPC1L1 knock-out mice to wild type mice fed on a high fat diet indicates that the absence of NPC1L1 is protective against obesity and type II diabetes. In addition, it has been found that NPC1L1 is highly expressed in white adipose tissue, which is involved in the development of obesity as well as diabetes. Thus, inhibitors of NPC1L1 would be capable of treating obesity and diabetes in a subject, in addition to hyperlipidemia and other lipid-related disorders such as cardiovascular disease. Several inhibitors of NPC1L1 have been identified, as set forth above. In addition, a transgenic mouse that overexpresses NPC1L1 has been created. This transgenic animal is useful for the identification and validation of agents that modulate NPC1L1.

* * *

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

It is further to be understood that all values are approximate, and are provided for description.

Patents, patent applications, publications, product descriptions, Accession Nos., and protocols are cited throughout this application, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

WHAT IS CLAIMED:

1. An isolated nucleic acid encoding a Niemann-Pick C1-like protein (NPC1L1) wherein the nucleic acid comprises a nucleotide sequence that hybridizes under normal conditions to the complement of the nucleotide sequence set forth in SEQ ID NO: 2.

2. An isolated nucleic acid encoding a NPC1L1 polypeptide, wherein the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO: 2.

3. An isolated NPC1L1 nucleic acid comprising a nucleotide sequence having at least 95% identity with the nucleotide sequence set forth in SEQ ID NO: 2.

4. An isolated nucleic acid comprising a nucleotide sequence encoding an NPC1L1 polypeptide having an amino acid sequence set forth in SEQ ID NO: 3.

5. An isolated nucleic acid comprising a nucleotide sequence encoding an NPC1L1 polypeptide having an amino acid sequence having at least 95% identity with the amino acid sequence set forth in SEQ ID NO: 3, wherein the encoded polypeptide has a lipid permease function.

6. An isolated NPC1L1 polypeptide comprising an amino acid sequence encoded by the nucleic acid sequence of claim 1.

7. An isolated NPC1L1 polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3.

8. An isolated NPC1L1 polypeptide comprising an amino acid sequence having at least 95% identity with the amino acid sequence set forth in SEQ ID NO: 3, wherein the NPC1L1 polypeptide has a lipid permease function.

9. A vector comprising the NPC1L1 nucleic acid of claim 1.

10. A vector comprising the NPC1L1 nucleic acid of claim 2.

11. A host cell that has been engineered to contain the vector of claim 9.

12. A host cell that has been engineered to contain the vector of claim 10.

13. An antibody that specifically binds to the NPC1L1 polypeptide encoded by a nucleic acid of claim 1.

14. The antibody of claim 13, which specifically binds to the NPC1L1 polypeptide of claim 7.

15. The isolated nucleic acid of claim 2 comprising a mutation in at least one nucleotide that results in defective expression or activity of the NPC1L1 protein product.

16. The isolated nucleic acid of claim 15, wherein defective expression of NPC1L1 results in a disorder in glucose metabolism.

17. The isolated nucleic acid of claim 15, wherein defective expression of NPC1L1 results in a disorder in lipid metabolism.

18. The isolated nucleic acid of claim 17, wherein the lipid is selected from the group consisting of cholesterol, triglycerides, and sphingolipids.

19. The isolated nucleic acid of claim 18, wherein the lipid is cholesterol.

20. A method of inhibiting the uptake of a lipid by a cell or transport of a lipid by a cell comprising contacting the cell with an agent which inhibits NPC1L1 nucleic acid expression or NPC1L1 polypeptide activity.

21. The method of claim 20, wherein the lipid is selected from the group consisting of cholesterol, oleic acid, and sphingolipid.

22. The method of claim 20, wherein the lipid is cholesterol.

23. A method of decreasing the plasma glucose of a subject in need of such treatment which comprises administering to the subject a therapeutically effective amount of an agent which inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

24. A method of treating hyperlipidemia in a subject comprising administering to the subject a therapeutically effective amount of an agent which inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

25. A method of treating type II diabetes in a subject comprising administering to the subject a therapeutically effective amount of an agent which inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

26. A method of treating obesity in a subject comprising administering to the subject a therapeutically effective amount of an agent which inhibits the expression or activity of an NPC1L1 nucleic acid or polypeptide.

27. The method of any one of claims 20, 23, 24, 25, or 26, wherein the agent is an antisense molecule or an siRNA molecule specific for an NPC1L1 nucleic acid.

28. The method of claim 27, wherein the siRNA comprises any one of SEQ ID NO: 23 or SEQ ID NO: 24.

29. The method of any one of claims 20, 23, 24, 25, or 26, wherein the agent is an antibody specific for an NPC1L1 polypeptide.

30. The method of any one of claims 20, 23, 24, 25, or 26, wherein the agent is a small molecule.

31. The method of any one of claims 20, 23, 24, 25, or 26, wherein the agent is a molecule selected from the group consisting of: 4-phenyl-4-piperidinecarbonitrile hydrochloride, 1-butyl-N-(2,6-dimethylphenyl)-2 piperidinecarboxamide, 1-(1-naphthylmethyl)piperazine, 3{1-[(2-methylphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3-[(4-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, 3-[(2-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, and N-(4-acetylphenyl)-2-thiophenecarboxamide.

32. The method of claim 24, wherein the hyperlipidemia is dietary hypercholesterolemia.

33. A method for identifying a test compound that binds to an NPC1L1 polypeptide, which method comprises:

(i) contacting a host cell that expresses an NPC1L1 polypeptide with a test compound; and

(ii) identifying a test compound that binds to said host cell but not to a control cell that does not express NPC1L1 polypeptide.

34. A method for identifying a test compound that modulates the activity of an NPC1L1 polypeptide, which method comprises:

(i) providing a host cell that expresses a functional NPC1L1 polypeptide,

(ii) contacting said host cell with a test compound under conditions that would otherwise activate the activity of said functional NPC1L1 polypeptide; and

(iii) determining whether said host cell contacted with said test compound exhibits a modulation in activity of said functional NPC1L1 polypeptide.

35. A method for identifying an agent useful in the prevention or treatment of an NPC1L1-mediated disease or disorder, which method comprises determining the effect of the substance on a biological activity of an NPC1L1 polypeptide by:

(a) contacting a test cell which expresses a functional NPC1L1 polypeptide with the test agent in the presence of extracellular cholesterol under conditions where uptake of the cholesterol would be effected; and

(b) observing the effect of the addition of the agent on the test cell, in comparison with the effect of a control cell expressing a functional NPC1L1

polypeptide not contacted with the test agent, wherein inhibition of cholesterol uptake in the test cell compared to the control cell is indicative that the test agent is useful for the treatment of an NPC1L1-mediated disease or disorder.

36. A non-human animal which has been engineered to be deficient in the expression of a functional NPC1L1, wherein the non-human animal does not express an NPC1L1 nucleic acid or polypeptide.

37. The non-human animal of claim 36, wherein said non-human animal is a mouse.

38. A genetically modified, non-human animal comprising a recombinant nucleic acid molecule containing a nucleic acid encoding an NPC1L1 gene product, wherein said animal has increased NPC1L1 expression or activity, or displays symptoms of hyperlipidemia, obesity, diabetes, or cardiovascular disease.

39. The non-human animal of claim 38, wherein said non-human animal is a mouse.

40. A method of screening for an agent capable of treating an NPC1L1-mediated disease or disorder comprising administering to the non-human animal of claim 33 a candidate compound and monitoring the expression or activity of NPC1L1.

41. A method of assessing whether a patient is afflicted with an NPC1L1-mediated disease or disorder or at risk for developing an NPC1L1-mediated disease or disorder, the method comprising comparing: a) the level of expression or activity of an NPC1L1 nucleic acid or polypeptide in a patient sample, and b) the normal level of expression or activity of the NPC1L1 nucleic acid or polypeptide in a control sample derived from a subject not afflicted with the NPC1L1-mediated disease or disorder, wherein a significant increase in the level of expression or activity of the NPC1L1 nucleic acid or polypeptide in the patient sample is an indication that the patient is afflicted with an NPC1L1-mediated disease or disorder or at risk for developing an NPC1L1-mediated disease or disorder.

42. The method of claim 41, wherein the NPC1L1-mediated disease or disorder is selected from the group consisting of hyperlipidemia, obesity, type II diabetes, and cardiovascular disease.

43. A method for inhibiting the expression or activity of an NPC1L1 molecule comprising contacting an NPC1L1 molecule with an agent selected from the group consisting of: 4-phenyl-4-piperidinecarbonitrile hydrochloride, 1-butyl-N-(2,6-dimethylphenyl)-2 piperidinecarboxamide, 1-(1-naphthylmethyl)piperazine, 3{1-[(2-methylphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3-[(4-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, 3-[(2-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, and N-(4-acetylphenyl)-2-thiophenecarboxamide.

44. A method for inhibiting the expression or activity of an NPC1L1 molecule comprising contacting a cell expressing an NPC1L1 molecule with an agent selected from the group consisting of 4-phenyl-4-piperidinecarbonitrile hydrochloride, 1-butyl-N-(2,6-dimethylphenyl)-2 piperidinecarboxamide, 1-(1-naphthylmethyl)piperazine, 3{1-[(2-methylphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 3{1-[(2-hydroxyphenyl)amino]ethylidene}-2,4(3H, 5H)-thiophenedione, 2-acetyl-3-[(2-methylphenyl)amino]-2-cyclopenten-1-one, 3-[(4-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, 3-[(2-methoxyphenyl)amino]-2-methyl-2-cyclopenten-1-one, and N-(4-acetylphenyl)-2-thiophenecarboxamide.

45. A method for inhibiting the expression or activity of an NPC1 molecule comprising contacting a cell expressing an NPC1L1 molecule with 4-butyryl-4-phenylpiperdine hydrochloride.

46. The method of any one of claims 23, 24, 25 or 26, whereing the subject is a human.

FIGURE 1A

FIGURE 1B

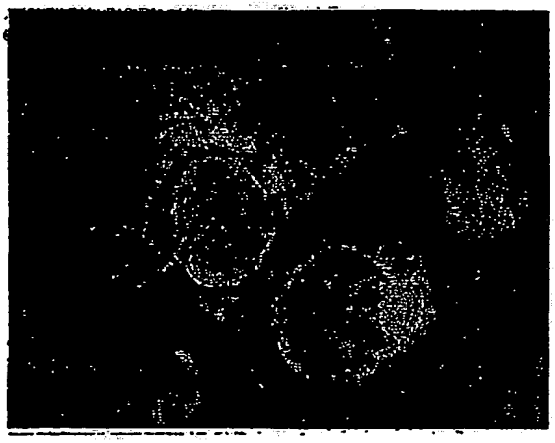


FIGURE 1C

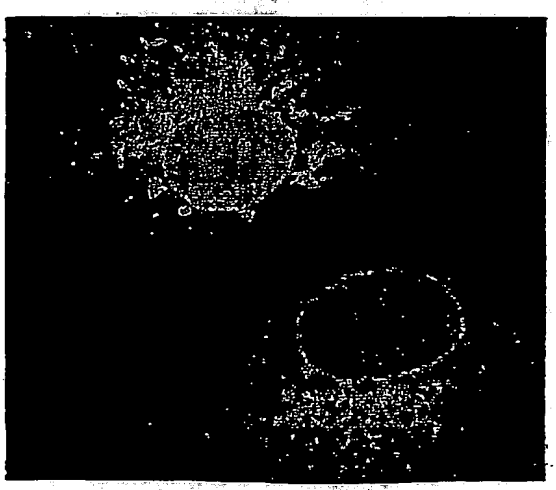


FIGURE 1D

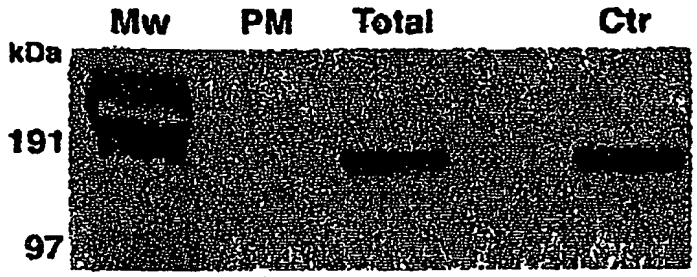


FIGURE 1E

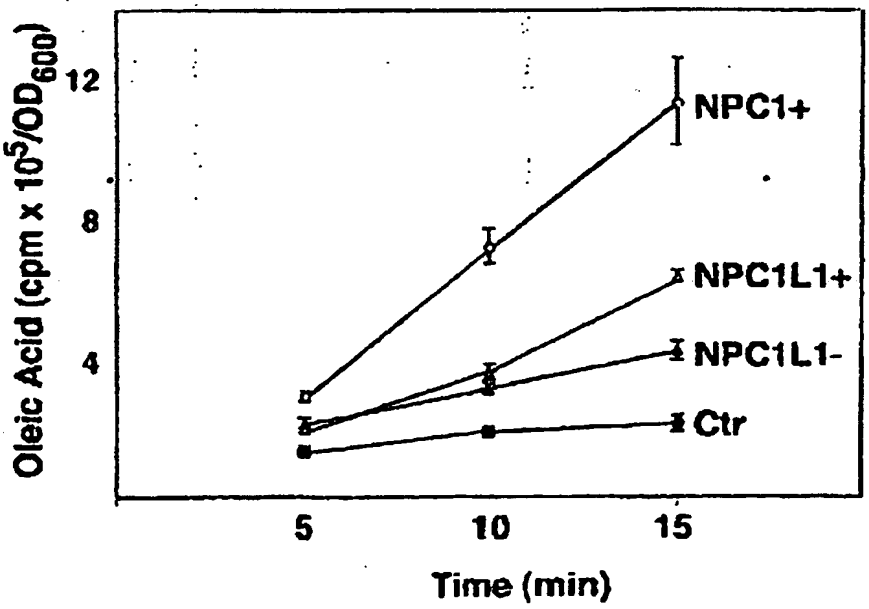


FIGURE 2A

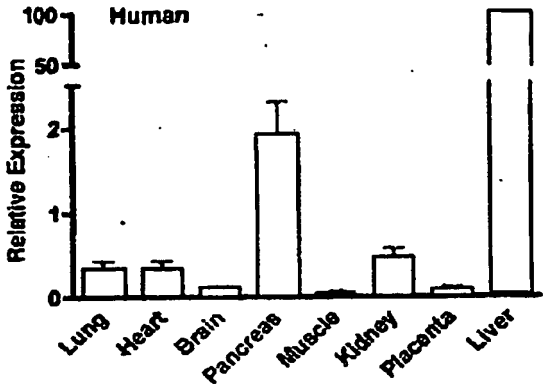


FIGURE 2B

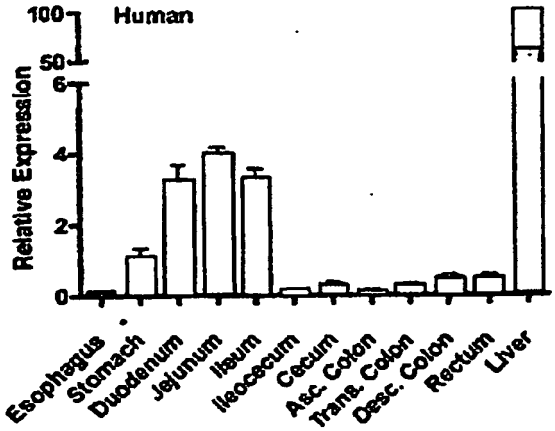


FIGURE 2C

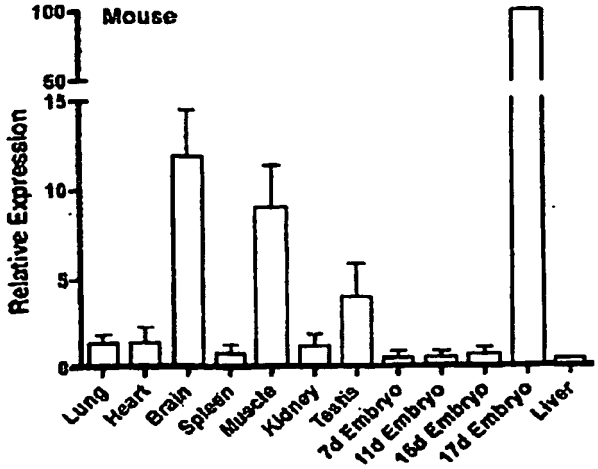


FIGURE 2D

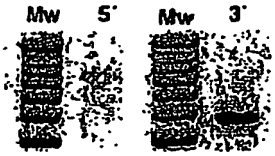


FIGURE 2E

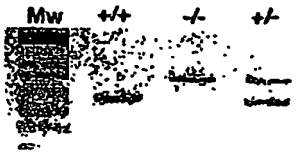


FIGURE 2F

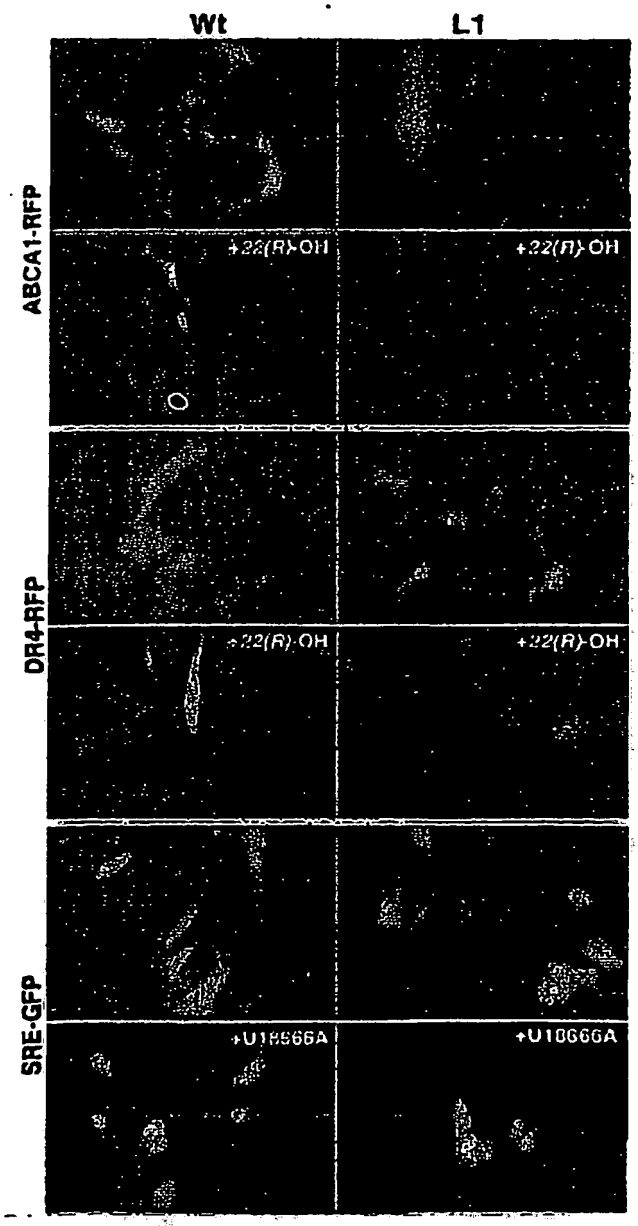


FIGURE 3A

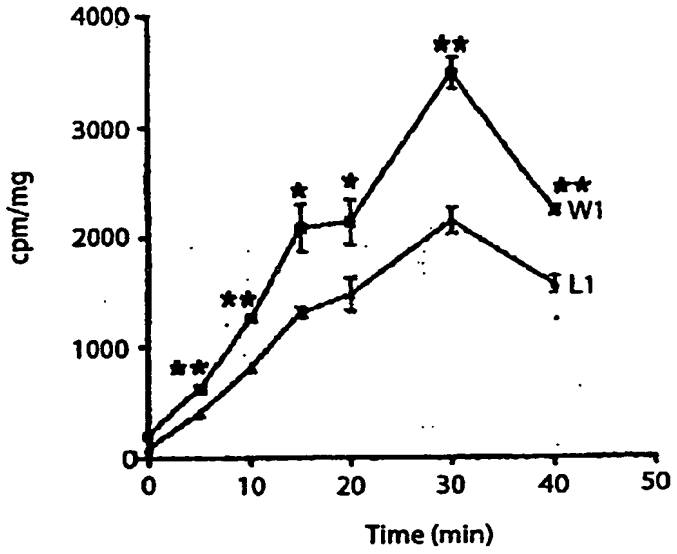


FIGURE 3B

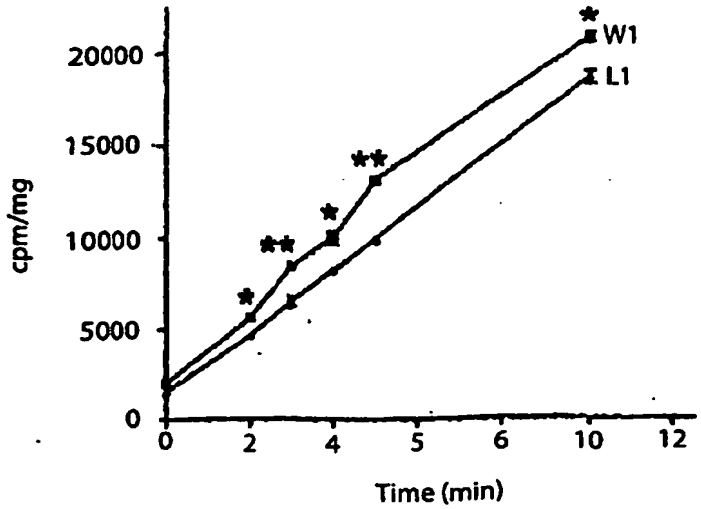


FIGURE 3C

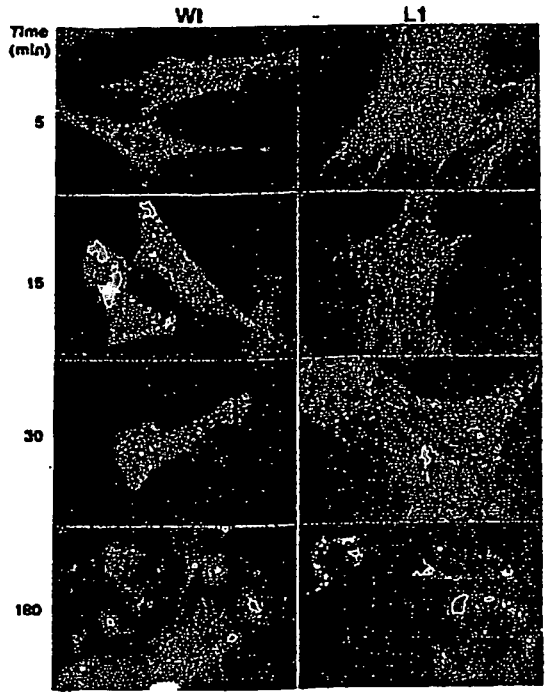


FIGURE 3D

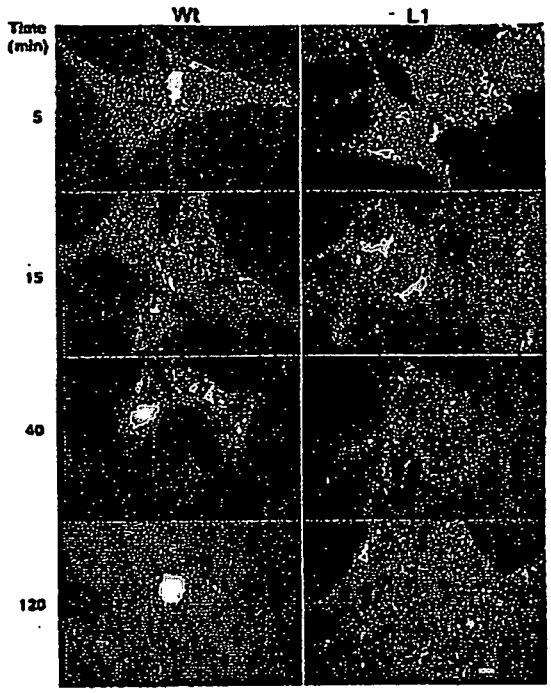


FIGURE 3E

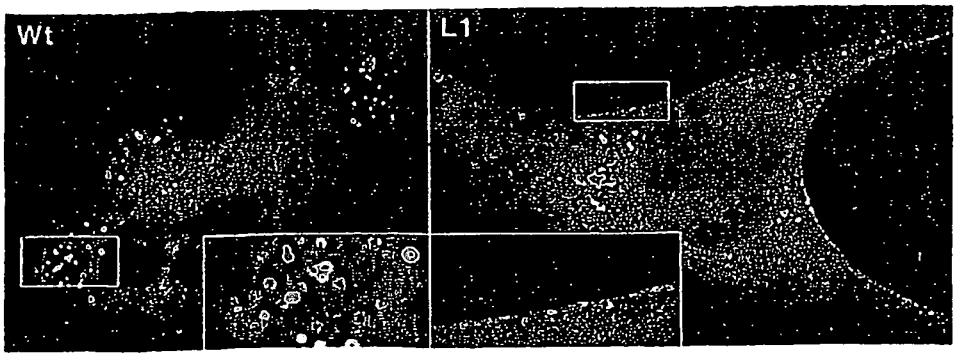


FIGURE 4

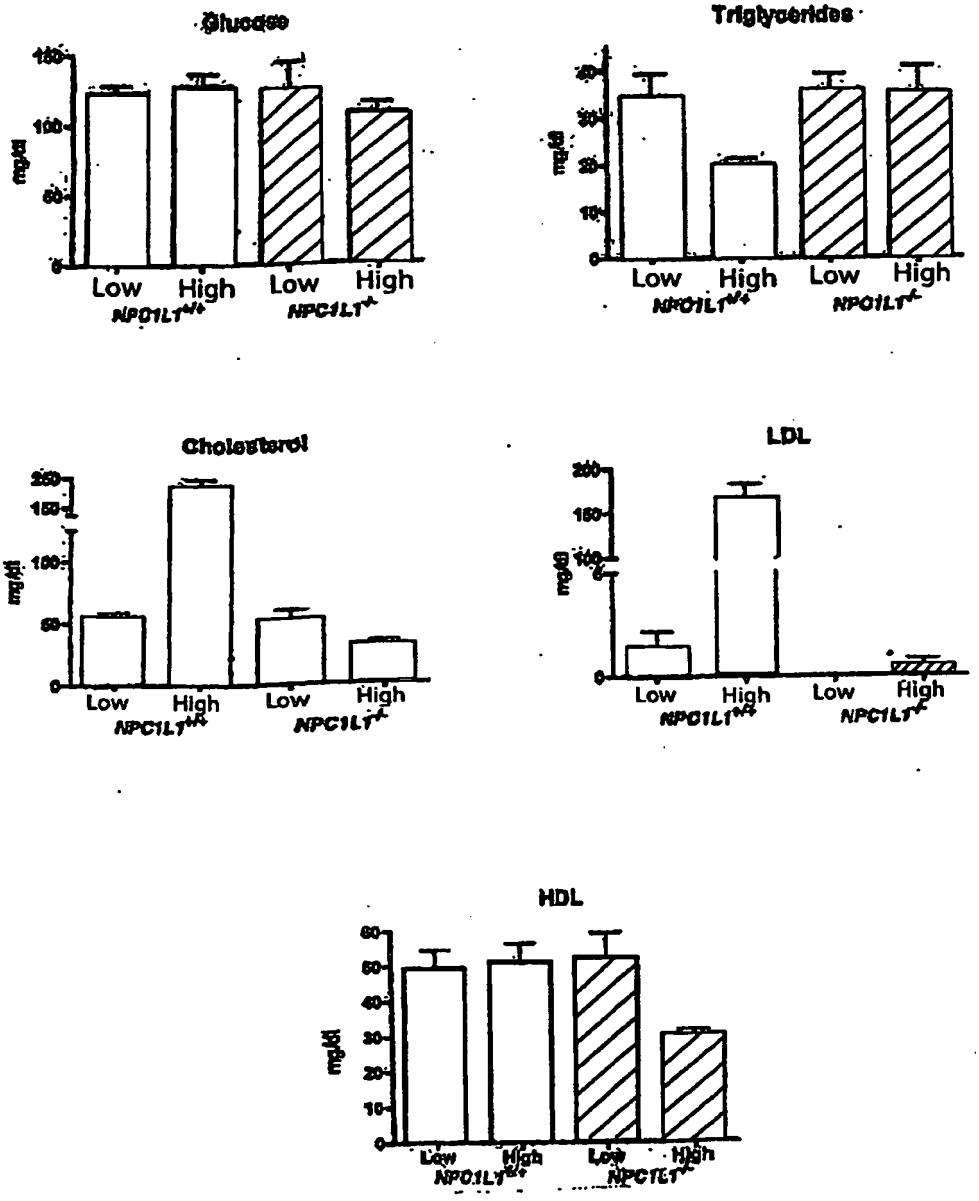

Glucose
mg/dl
Low
High
Low
High
NPC1L1+/+
NPC1L1-/-
Triglycerides
mg/dl
Low
High
Low
High
NPC1L1+/+
NPC1L1-/-
Cholesterol
mg/dl
Low
High
Low
High
NPC1L1+/+
NPC1L1-/-
LDL
mg/dl
Low
High
Low
High
NPC1L1+/+
NPC1L1-/-
HDL
mg/dl
Low
High
Low
High
NPC1L1+/+
NPC1L1-/-

FIGURE 5

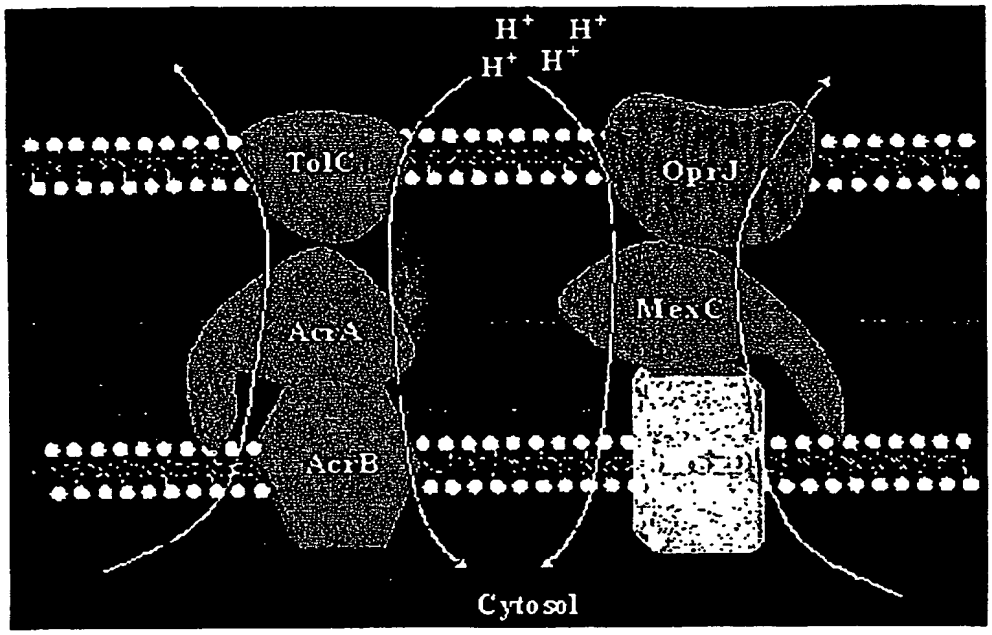


FIGURE 6A

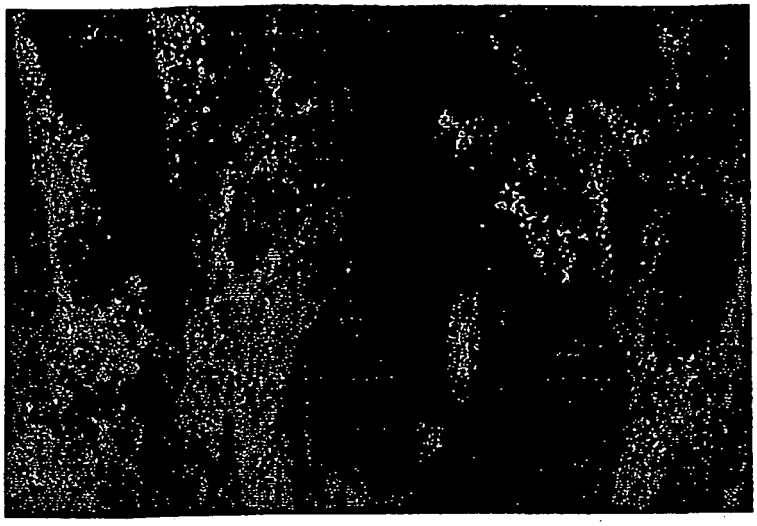


FIGURE 6B

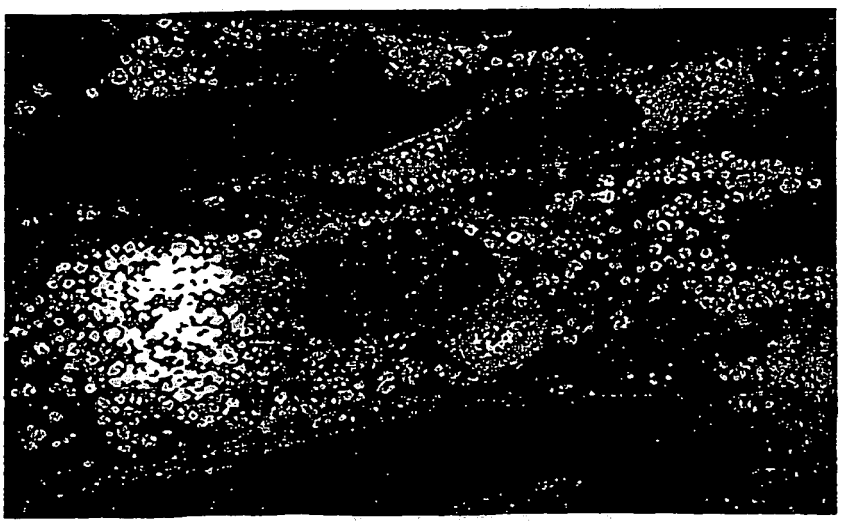


FIGURE 7A

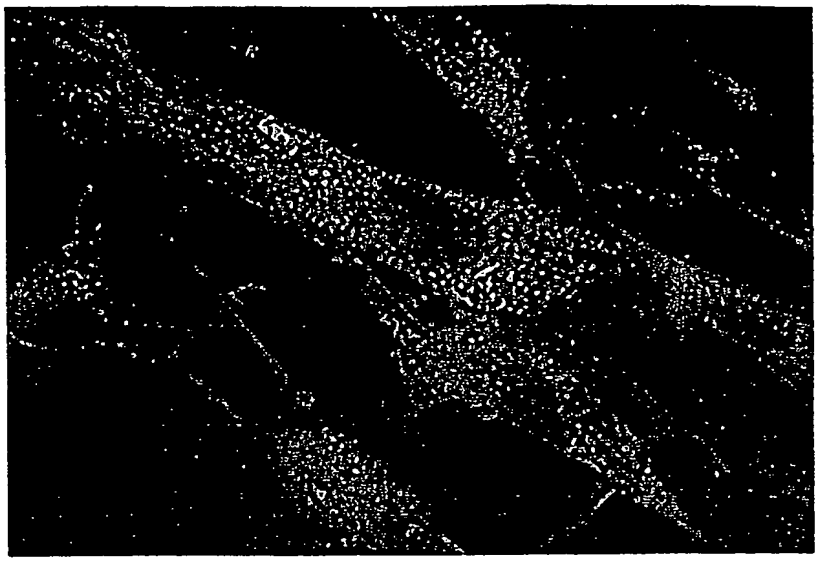


FIGURE 7B

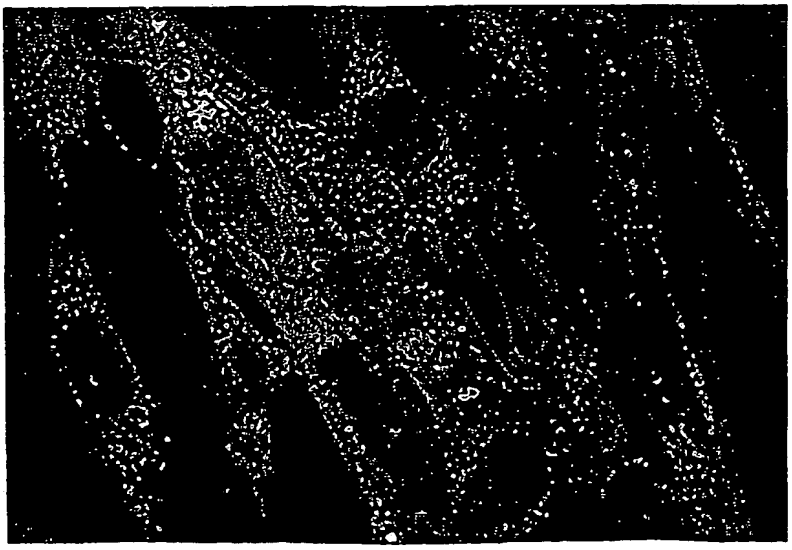


FIGURE 8

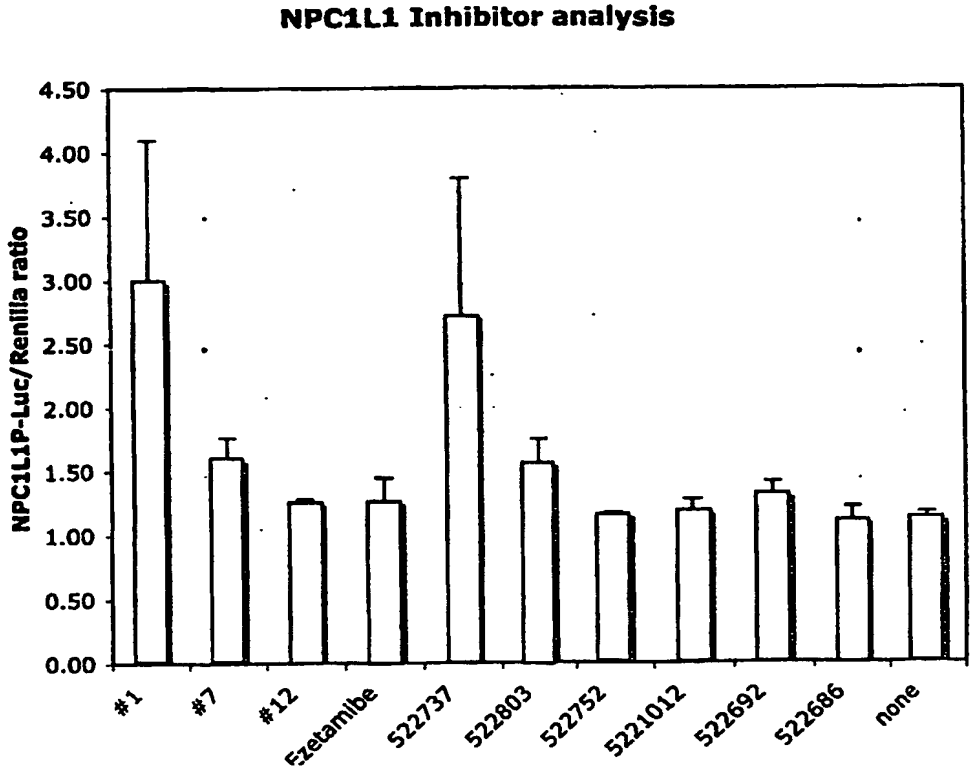

NPC1L1 Inhibitor analysis
NPC1L1P-Luc/Renilla ratio
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
#1
#7
#12
Ezetamibe
522737
522803
522752
5221012
522692
522686
none

FIGURE 9A

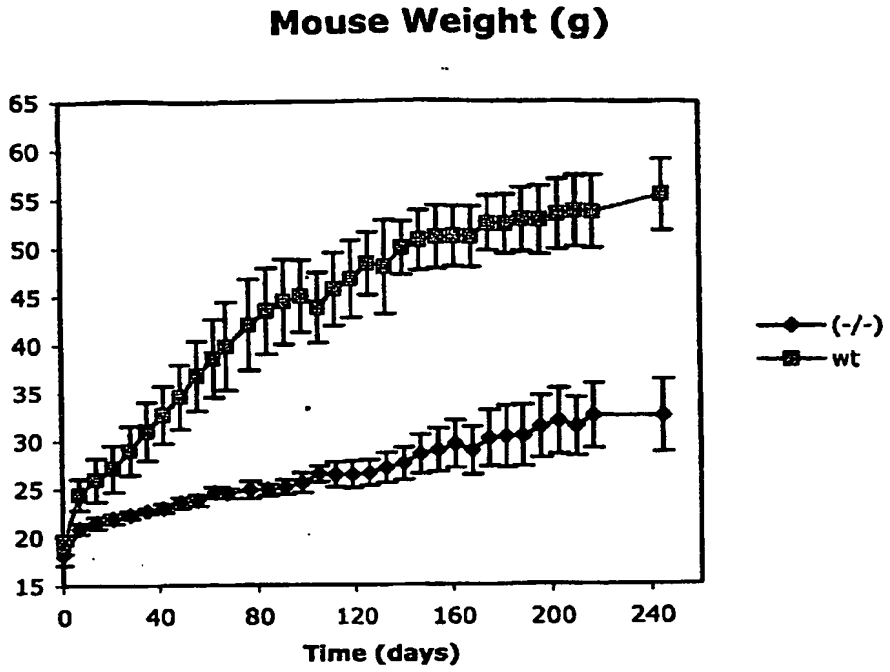


FIGURE 9B

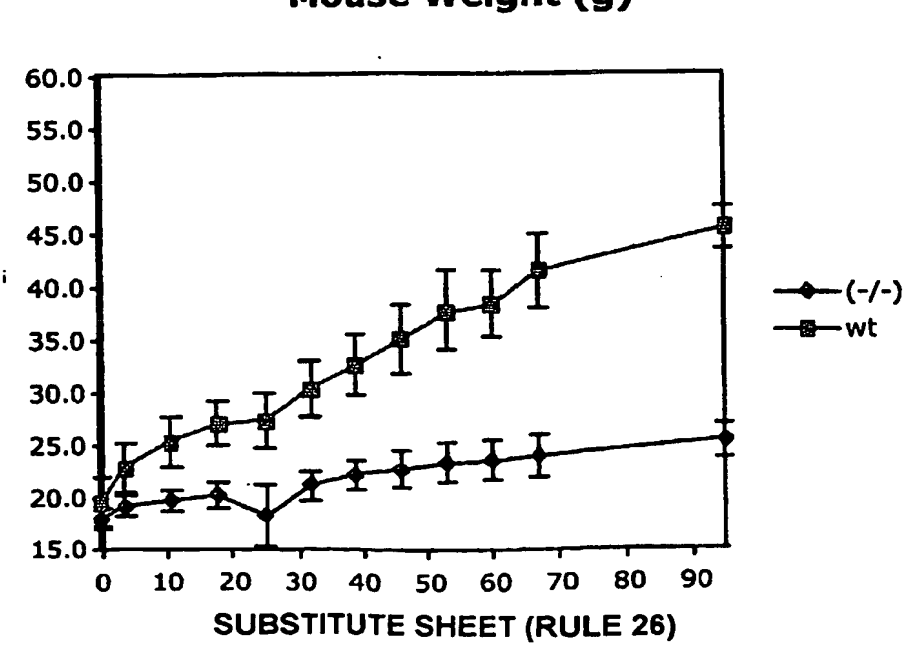


FIGURE 10

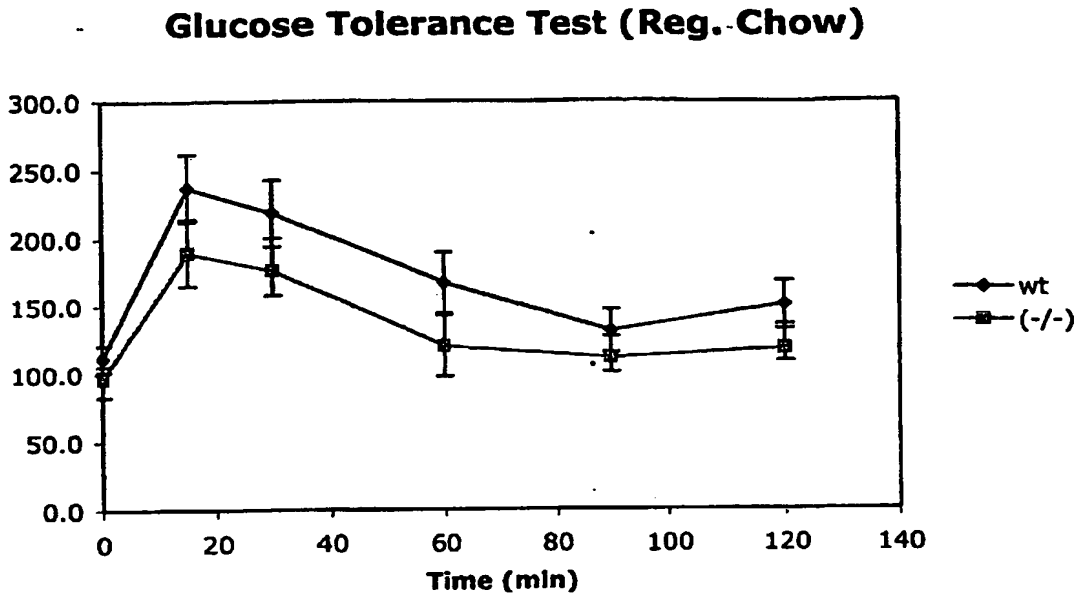

Glucose Tolerance Test (Reg.-Chow)
300.0
250.0
200.0
150.0
100.0
50.0
0.0
0
20
40
60
80
100
120
140
Time (min)
wt
(-/-)

FIGURE 11A

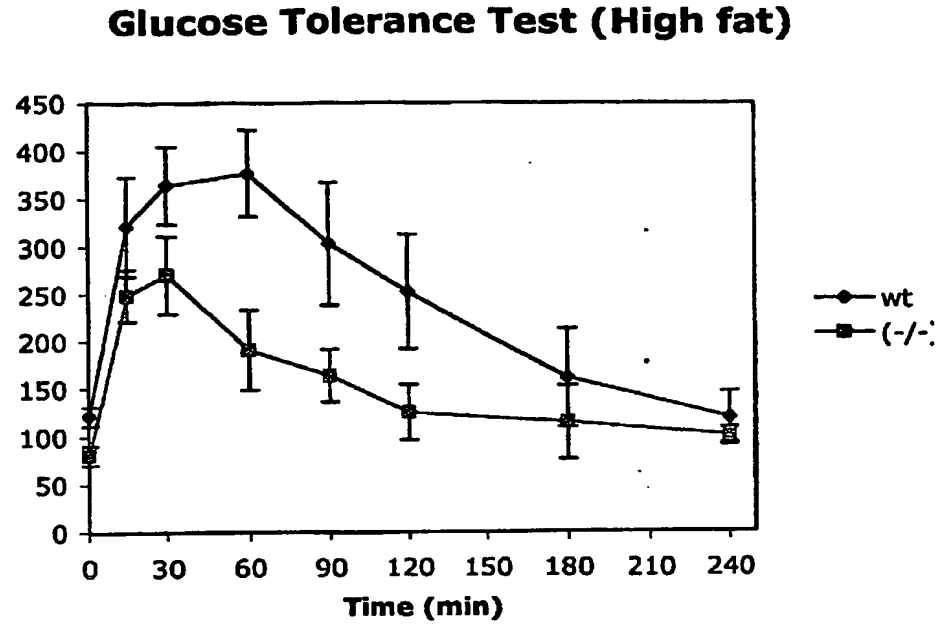


FIGURE 11B

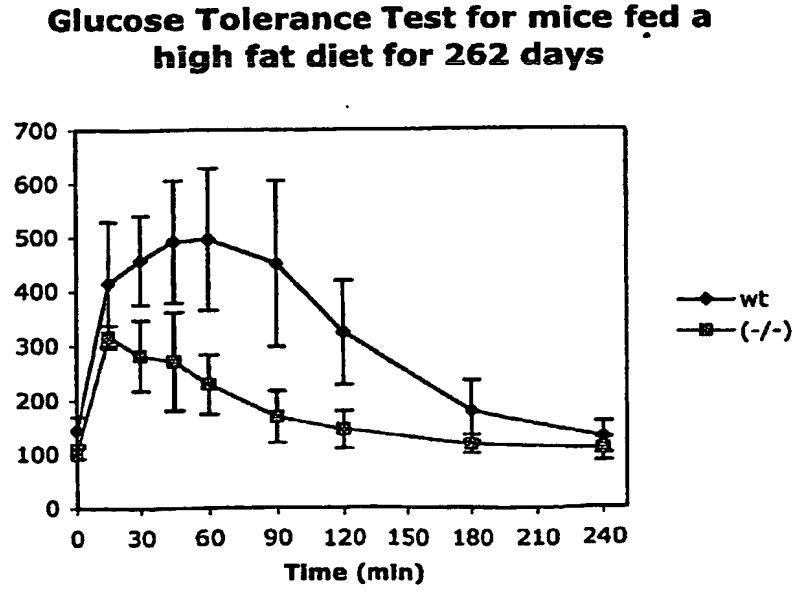


FIGURE 12A

Insulin Tolerance Test in mice fed a high Fat diet for 105 days

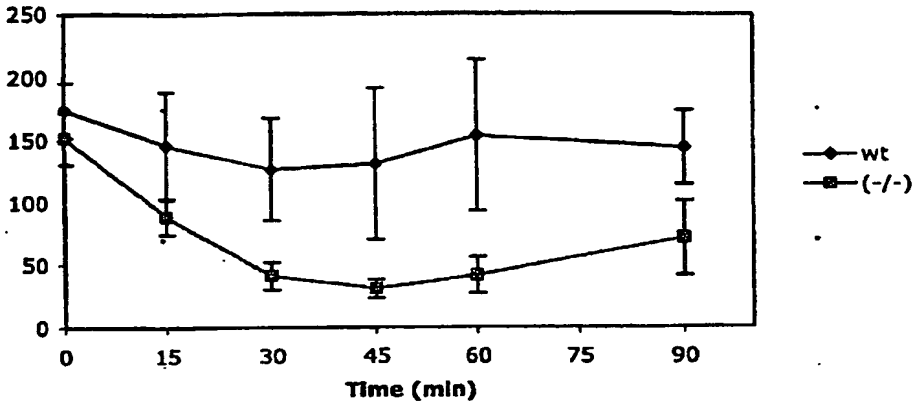


FIGURE 12B

Insulin Tolerance Test in mice fed a high fat for 252 days

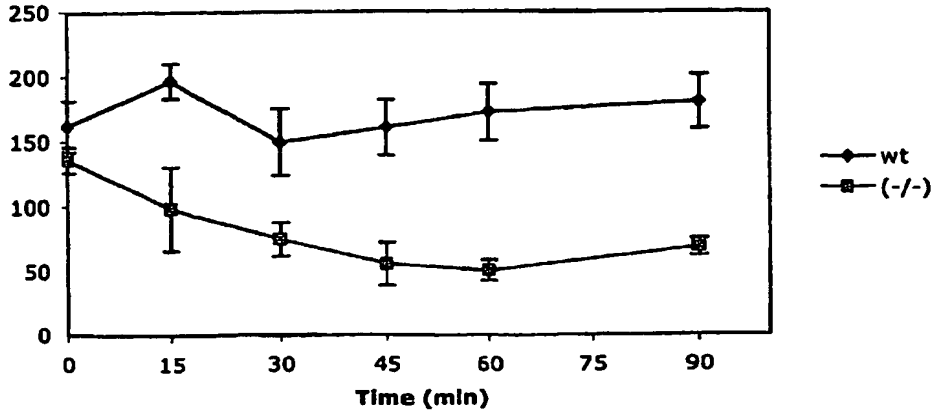


FIGURE 13A

Insulin Measurements in Mice fed a high fat diet for 72 days

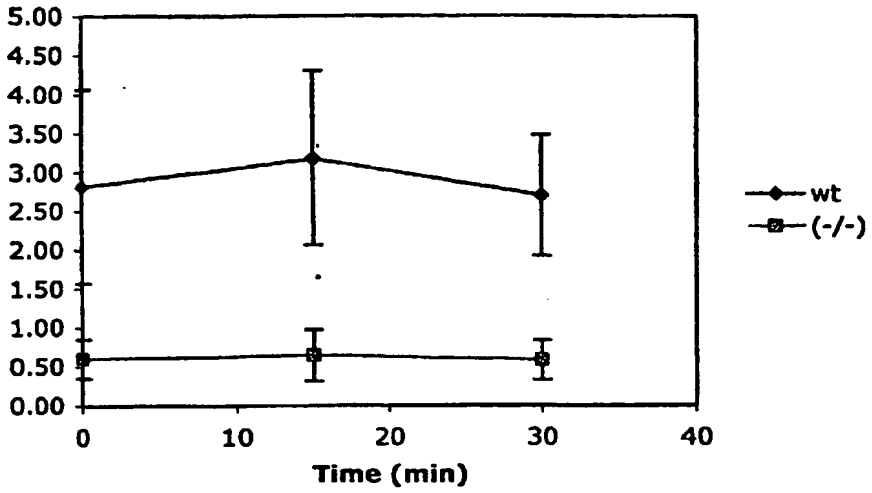


FIGURE 13B

Insulin Measurements in Mice on high fat diet for 220 days

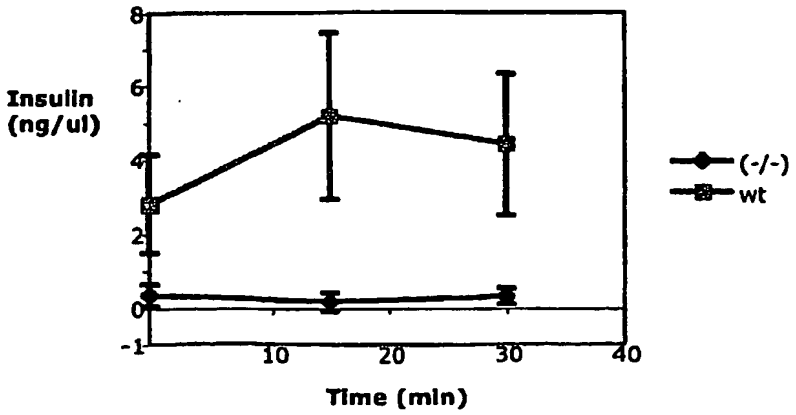


FIGURE 14A

Wildtype and NPC1L1-deficient mice fed High Fat diet for 120 days

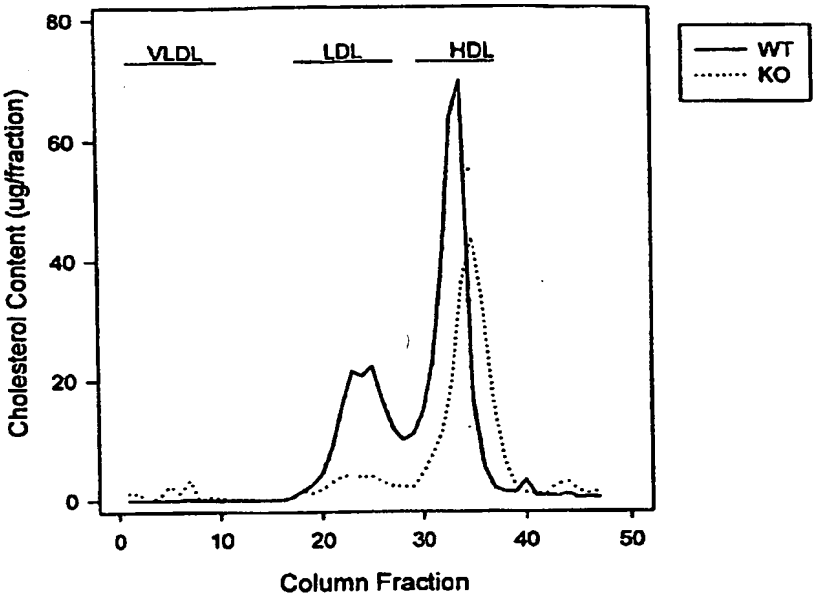


FIGURE 14B

Wildtype and NPC1L1-deficient mice fed High Fat diet for 268 days

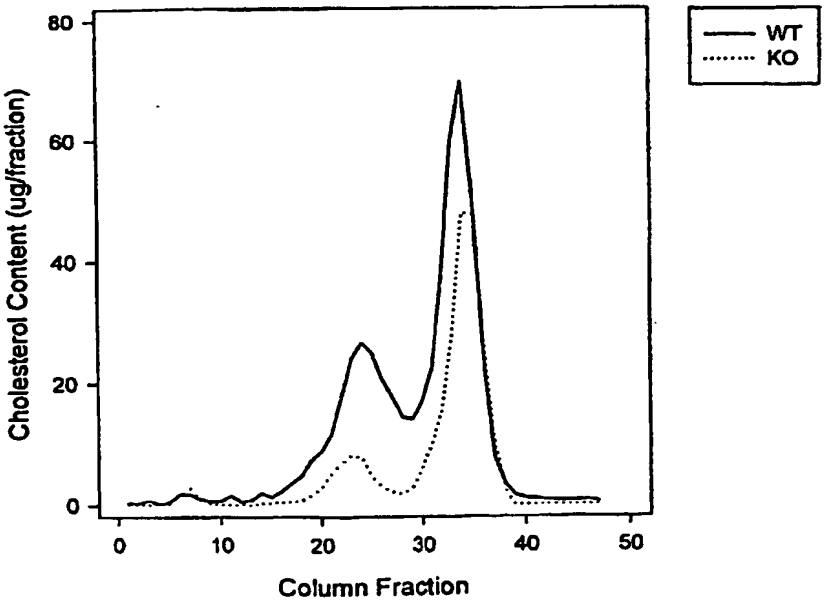


FIGURE 15

NPC1L1/β-Actin Expression in Mouse tissues

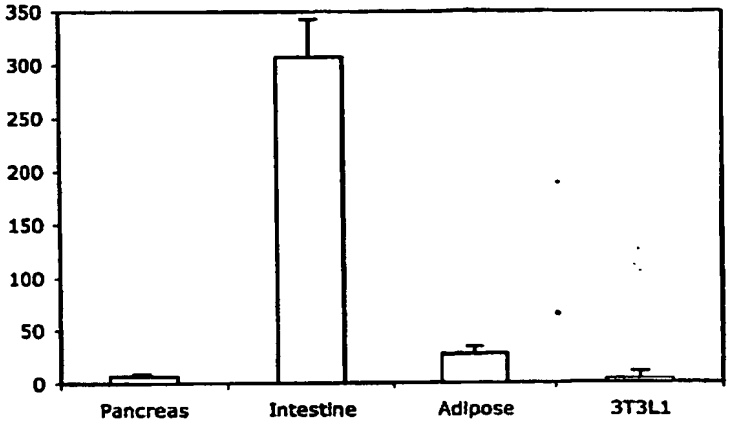


FIGURE 16

NPC1L1/β-Actin Expression in Mouse tissues

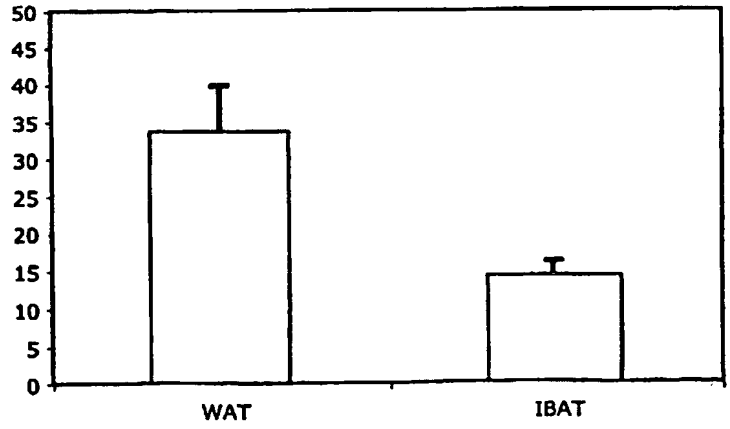


FIGURE 17

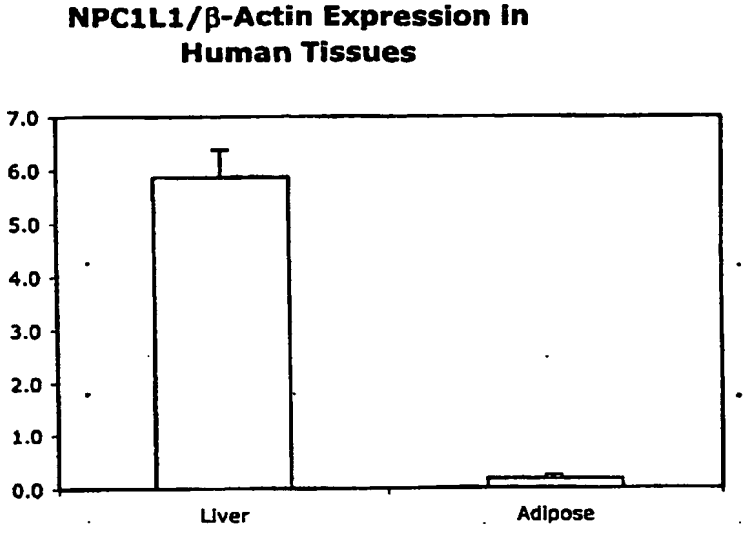

NPC1L1/β-Actin Expression in Human Tissues
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Liver
Adipose

FIGURE 18

Days on high fat diet		35	42	49	56	63	84	91	98	105	112	119
g food/day	(-/-)	2.70571	1.24	1.72571	1.76571	1.58857	1.71429	1.71143	1.81143	1.74286	2.38857	0.93429
g food/day	wt	6.25469	4.84286	6.12857	5.14898	3.54898	3.48367	3.28776	3.7449	2.91224	4.56327	1.11429
Weight	(-/-)	22.68	23	23.54	23.78	24.6	24.84	25.16	25.6	26.54	26.42	26.44
weight	wt	30.9571	32.6714	34.5429	36.7286	38.5714	43.5	44.4714	45.0571	43.8571	45.7571	46.8
food/d/g wt*10	(-/-)	1.193	0.53913	0.7331	0.74252	0.64576	0.69013	0.68022	0.70759	0.65669	0.90408	0.35336
food/d/g wt*10	wt	2.02044	1.48229	1.77419	1.4019	0.92011	0.80084	0.7393	0.83114	0.66403	0.99728	0.2381
% intake by NPC1L1 of wt	(-/-)	59.0464	36.3714	41.3201	52.9653	70.1833	86.1754	92.0089	85.1344	98.8948	90.6543	148.411

Wt n=7
L1 (-/-) n=6

(continued)

Days on high fat diet	(-/-) wt	126	133	133	140	147	154	161	168	189	196	203	210
g food/day		1.82571	1.83714	1.83714	1.79714	1.93143	1.79429	1.89429	1.76857	1.88857	1.97143	1.98	1.68857
g food/day	(-/-) wt	3.11429	3.43469	3.43469	2.9551	3.24694	2.87857	2.93095	2.72381	3.05238	2.51429	2.99762	3.01667
Weight		26.58	27.08	27.08	27.54	28.5	28.92	29.52	28.88	30.42	31.36	31.92	31.9
weight	(-/-) wt	48.3571	47.9571	47.9571	49.9143	50.7333	51.05	51.15	51.1	52.9333	52.85	53.45	53.7667
food/d/g wt*10		0.68688	0.67841	0.67841	0.65256	0.67769	0.62043	0.6417	0.61239	0.62083	0.62864	0.6203	0.53176
food/d/g wt*10	(-/-)	0.64402	0.7162	0.7162	0.59204	0.64	0.56387	0.57301	0.53304	0.57665	0.47574	0.56083	0.56107
% intake by NPC1L1 of wt	(-/-)	106.655	94.7239	94.7239	110.223	105.89	110.03	111.987	114.887	107.663	132.14	110.605	95.8463

Wt n=7
L1 (-/-) n=6

SEQUENCE LISTING

<110> Mount Sinai School of Medicine
Ioannou, Yiannis
Davies, Joanna P.

<120> NPC1L1 AND NPC1L1 INHIBITORS AND METHODS OF USE THEREOF

<130> 2201581-WO0

<150> 60/592,592
<151> 2004-07-30

<160> 33

<170> PatentIn version 3.3

<210> 1
<211> 50000
<212> DNA
<213> Mus musculus

<400> 1
cttttttgac agccaaatct ttttttattg ggggaacggg tctctagggg gtaggcctag 60

gccctcactg cacagcttgt tcattggcac tgcctccaga atcctgtggc ttcatcacat 120

ctggaagctc gggagggctg gagaagggct caatgcggag agtttcgaag gtgtcatctt 180

ctcggaaggc caggcccact gtggctgtgc tgtctggcta gtgaagccac actcgcccag 240

agttttgcca tcatcaagga gctggtcatc cttgtaaagc agccgctcct ctggcggccg 300

cttgaggatg tcctcgacga tgtgcttcaa ttcgaacaca gtgctcaact tcttggcatc 360

cagaaagatg gtggtcttgt ggcaccggat catgagaaac atgtccattc tggcggctgc 420

ttctggcttg aggcgccagt gcagccccaa ttgtggcttt ctttgtttct ttttttttta 480

aagctttatt tatttattaa ttatatgtaa gtacactgta gttgtcttca gacaccccag 540

aaaaaggtaa catctcatta cggatggttg tgagcaacca tgtggttgtt gggatttgaa 600

ctcaggactt ccagaaaagc agtcagtgct cttaagtgct gagccatttc tccagcccaa 660

ttgtagcttt ctataatggt gtctgtctgt agcaaagaaa ggcttctttt tattgttatt 720

atatattgta tatataaaat ttttcttttt attgctatta tgtattggat atataggatt 780

gtatatatat gcatatgtag ttatatattt atatattaca tacatacata tatatagttg 840

ttataatttt attttatgtt tatggatgtt ttgcctgcat gtatgtctgt accgtgtgtg 900

tgtgtgtgtg cgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtatctg gtgcctgagg 960

aagtgagaag agggtgtcag atcccctgga actgtagtta tatgttgcgc gcgctcaact 1020

ggccaggaag aacgacgctg ctacaggatc cttctgcaca catttattca gtcctgtttc 1080

ttctttctcc atatatctcc cttgtttata tctcccttgt ttatatctcc cttgtttata 1140

tctctccctt gtttatattt cccctgtttg tatctctccc tcgtttatat ctcccccttg 1200

tttatatttc ccctgtttgt atctctccct cgtttatatc tcccccgaac cctgggcctc 1260
tcactctttt tatactctct ctcatccacg cactgcaggc cacgcccccc cgccagtcac 1320
gaggcttcag ctaatcaggg cagcaggggc aaatctccac caaattggat tcacctgtat 1380
cctggtacac ctgcgcagca ctcaagatgt ttgtgtctta tatgaggaag tcaggtgcaa 1440
gtcatatgac ttagctgcag tccctggcgc ctttaggact gccgccacac ctgctcctaa 1500
caattccccc ttttttcttt tttggcagag agaatgcctg tagatagccc cccgcagcca 1560
tgccccttac ccgtccttgg gtgacaaaca gcattggttt gatccctgtc ttaggttggt 1620
gacatgccca gggagtctta tcactgacta cctctctatc atgccaagcc acacctgggg 1680
agattgtgtt ttgcttgcgg caggtgcatc aaaaggccag gatgttaatt aacctatggc 1740
cagatcttaa ttgctgcaag caagcctcat gggtgaaggt agaggtctga tccccagtgt 1800
gcaagcatag gggccctaca caaaaccatc ccttaggctc atcacccaga gaggtcttgg 1860
ccagctcccg tgtcgttttt cctgggggaa gggaactagg acactgaacc ttcatgcaat 1920
cagacatgcc ttccacagga tgccaacagc aagctatcct ctgtgcagtg cttagcctcg 1980
caccccacgg cataacgcag cataatttct tttagagcgg caaaccgaat ctgaggagat 2040
tgtcccgcct ccactgcagc aaaagcctac gctaccatgg cgaaagtgcg ggtcgcacac 2100
tctgcaccta acacatgtgc caaacacaaa acacacacac actataacaa gcatacatcc 2160
cagggctctc atccccactc atcttccctg aagcaaggga tagatagcca gaggggctat 2220
ctctttaaga ggaattcagg gatcaaaaag cgtggaaaaa tttgaatgtc atgtcgagct 2280
ggtatcggct tctggaatac cgaaaggatc tcccatctct gctccatcct ttgtgcttgt 2340
gggatcatcc accacatcca caggggcaac tggatcagga gcctcattct tgcagtgtct 2400
caccaatctc tccggcaccc aaagaggttc catctggtcc tgtggaaaca cacaaacaga 2460
ccctctcgcc ctcgtcaaca ccggatctgc tcgatcccca ggggagatgg acacaatacc 2520
tcgtggtaat gagaccataa cctcaatcac ctgtgtgttc atctggggag tcaatacgcc 2580
aaatctccgg acagaaggtc cacctatagg tggctgctga ggaggcatag ggaggaggca 2640
cagaagctaa ttcgccttcc tcctccacct cggctctttt attttgtcct ttctgtccac 2700
ctgataacac tgtgtctcct ttctttttcc tttttctttt taagcccttc tccttttccg 2760
acatactttc ttgttgctct gtaaggattt tctgacctgt cttgactgcc tctacacaca 2820
aacagtaaca gagtccatat atcagaacaa acaagaccaa aactatcaga cctacccaca 2880
aagggtcaat gggagaaaaa agcataacta cacagcgagg atctattgat cactacactg 2940
agcttatcat agttccttaa cttgtcccaa agccaggaac cttaacttgt cccaaagccg 3000
ggaacctttc gttaccttgt gcctgcttgc tggcaacttt atgctcacct ctattttatc 3060
agaggtcctt cccaaactcc tgggttactt tgtgcctact tcctggcaac tttatgcttg 3120

cctctatttt atccgaggtc cttcccaaac tcctggggtt gcaagtttca ctcaccgtga 3180
acttaccctg cctgccggca accactgact gctgaaagtt ctgaactcgg tgggggagtc 3240
ggttccccgt acgggccacc aattgtcgcg cccactctcg accagcaaga acgacgcgac 3300
caccagtcct tctaacagca gtttattcag tcttcatctc tcttctttct cttcatcagt 3360
accgttcccc agctgaagag ttctgaatcc acgccggatc cttctcaaca gtctgtttca 3420
cgggaacctt tattaaccgc tccttccccg tgatgcagtt ctgaatcctc cctgtagcag 3480
ggggtcttcg ctcatgcctg aagatgtttc ttttcccggg tttcggcacc aactgttgcg 3540
cgcgctcaac tggccaggaa gaacgacgct gctacaggat ccttctgcac acatttattc 3600
agtcctgttt cttctttctc catatatctc ccttgtttat atctcccttg tttatatctc 3660
ccttgtttat atctctccct tgtttatatt tcccctgttt gtatctctcc ctcgtttata 3720
tctcccccct gtttatattt cccctgtttg tatctctccc tcgtttatat ctcccccgaa 3780
ccctgggcct ctcactcttt ttatactctc tctcatccac gcactgcagg ccacgccccc 3840
tcgccagtca cgaggcttca gctaatcagg gcagcagggg caaatctcca ccaaattgga 3900
ttcacctgta tcctggtaca cctgcgcagc actcaagatg tttgtgtctt atatgaggaa 3960
gtcaggtgca agtcatatga cttagctgca gtccctggcg cctttaggac tgccgccaca 4020
cctgctccta acagttatag atggttgtga gcctttgagt ggggactggg aatcaaactt 4080
gtcctctaga agagtagcca ctgctcttag atacttagcc atctttctag cctaaagaat 4140
ttttttccc tttggctttt caatacacgg tttcttggtg tagccctggc tgtcgtgtcc 4200
tgtacacaca catgcacatg cacacgcaca catgcacagt catccatgat ggagaggggg 4260
actgagcccc gggcctgaga tgccaagcac acacactgtc attgaactgt acctttagtc 4320
actaaaaagc cctggtctga cagccactgt gccgctgggc atggtggtgc aggcctttga 4380
ctccagcact tggaactcag cgtctggctg acctctgtga gtttgctgca agcttggtca 4440
acatacagag ttccaggcca gccagggcta catagtgagt gtggctgtgt ctcaaatgta 4500
aacaaaaagg ccttacacaa ccaagtcaaa ctcaaccacc ctttcttact gttttggtgt 4560
aagtgacagg acctcacttt gtgagaagct gtcagctgtt gccctaataa ttaaggttga 4620
agtctatcat tgtctgggta gccttctggg ctccatgcta atggtgaact ttctctgtca 4680
gactctcttc ttaacctggg ggatccctgt gatggtttgt atgttcttgg ctttatctcc 4740
tgggattaaa ggcatgcact accttgcctg ggcctaagct tttcatagct gctgtgcctc 4800
aagatctcca tgtcaagatc taggtcagaa acttgtgtct tccagcctca agatctggat 4860
cacatgtgag ccctccaatt ctggattgta gttcattcca gatatagtca agttgacaac 4920
caggaatagt cattacaatc caacccttgt cttgtttgtt tttttgtttg tttgtttttg 4980

ttgttcttgt ttttttcggt ttttcgagac agggtttctc tgtgtagccc tggctgtcct 5040
ggaactcact ctgtagacca ggctggcctc gaactcagaa atccacctgc ctctgcctcc 5100
catgtgctgg gattaaaggc atgcaccacc actgcccagc taatccaacc cttgtcaatt 5160
tgacacaaat atctcatgtc cacatgaaac aataacaaga tcataaatat gcctaacatg 5220
atataactat tccttgtaca atcacaaaaa catttgtaaa attacagtgg ggcaatgtcc 5280
ctcgggaaca ttcttttagt atctcaactt aaatacagat tgatgttaaa aaaaaaatgg 5340
gagaaagcac aaatagctat acaaatgtgt tcttaacaat ataaaccaga agcattgata 5400
ttactttata atcctcattt ctgcaactgg tcatgtggtc ttagatggta tttataacta 5460
cctccctcta ctacccattc tgtattttct ccatcctctg caagcacctc agctggtctt 5520
cttggctctt ttcctggagg agtgacccat accttcaccc ctgatgggtc tgtgtccttt 5580
gtcatcctgc ttggattagg ctgttttgtt ttctattgac tttaatcaca ggacttggta 5640
gtactaggag acaccctaag ggatctcctg cactccagac ataatccctt ttaccttcat 5700
tgtggtagtt gatccaattt ccccatggta atctggatct atcacccctc ctaacactgt 5760
tattcttttc ttagcctgtt ggtttaaggg cattagaagc ccaaaatgac aaggggaag 5820
tttgagcttc cagttaaatg aaatgttagt tgtggctcct ggcaggaaca cactccactc 5880
ctggtaccaa tttgtattag tcaggatttg tcagagacca taccgtgaga aatcgagctt 5940
ctcacaatct cccaccctga caggccaaat ggcctcgtga taggagccgg ttgtcactcc 6000
ctcctccctg ttcccttcct ggcacctgag gctgtgaaag ctgaattata gtccccgctt 6060
ccctatctct tcctgactcc atgacatcca aggacatgag ttacacctga gcccggcctg 6120
acacctcaag gctgttaagg aggatctatg ttctggagat aagatgcaga gtgcccaccg 6180
cctggagcct gactttggcc cttatgtcag cagatgtcca cttgtttgtt ctttgttaaa 6240
ttcccccttg acccctccct attccccaag atgtatgctt taaaaccagg catctcagta 6300
taatagatgg agaccttgat aggcaccctt cttggtctcc gcttctcctt cccttcttcc 6360
cattttcttc caggtttgcg gtccctctca cgaataactg aatcctgcgg gacgggataa 6420
gtggcaccca acatgagggt gaggattgta tttccttcca gtggaggttc cagagaggtt 6480
tgtcacgacc ccaagaattt agaagtagtg aaggacccct tctgccgctc acggaagagt 6540
gagaagtcct tggtgagttg agtcatctcc acttcaggtt atgggaaata agctttctaa 6600
agaggcagcc ttcatcaaag gcttaaagat agctctcagg gaaagaggag tacgagttaa 6660
aaagaaagat ttgataaact tttattttca tagaccaggt atgtccatgg tttattatag 6720
atgaagcaga gatacgttgt aaaaaatggt gaaaggtagg tagagactta aatgataaac 6780
tagctaatga gggtcccgat gcggtccctg caactgtctt ttcttattga ggagtaaata 6840
tcagaagctg ccactgtgcc tcctctccct tctctggcag aattcccaga agaaggagat 6900

aaagaattag agtctgaaca tgagaaagag aaaataagtt tcagaaaagc agttatcccc 6960
tgtttgggac cttttagcaa aaaagaggga aaatgaaaat aagctatatc agagctctcc 7020
tagggacaga aggaaaatgg gagacgtcct gcttcctctc tggctcctat ggagatattc 7080
ctagttctgg ttaggatgtc tgggtccctt tgagacatca agttctattc cctgtctgtc 7140
tattgtctgg ttttggtgca ccaaaatgtg tccatatatc tgtctattta tgtttgcttt 7200
tgttttgttg tttgaatgat agttgtactg tgtttcatgt tgaaaaaaat ggttaaaatt 7260
ttatctgctg gctgtccatt ctttgattta gtttaacttg tttaaaaagc agtctcaatt 7320
tgcacagcag aaacagataa ctgtggctgg gagtcaagta tagctggaag agccctgctg 7380
agagccggga cgggataagg attctctaga gtcacagaat ttatggtatg tctttctata 7440
ttaagggaat ttgttgtgat gacttacagt ctgtggtata gctaccccaa caatggtcag 7500
ctgtgaatgg gaagtccaag aatttagtag ttgcccactc cctcaaggtt agtgaggcta 7560
gttgttgtag ctgatcttct gtagaagtag attccaacag atgtgttggc aagtaaatgc 7620
aagcaggtga aggagagcaa atcttccttc ttccaatgtc tttacgtagg tctccagcag 7680
aaggtgtggc ccagattaaa ggtgtgtacc accatgcctg gatgggattt gttttatcct 7740
aggatgatct tgaactcaga gatctccttg ctttagtctc ctgggattaa gggcgtgtac 7800
taccttgcct gggcctaagc tttgctttgc tttgcttttt tttctttttt ctttttttctt 7860
ttggtttttc gagacagggt ttctctgtgt agccctggct gtcctggaac tcactctgta 7920
gaccaggctg gcctcgaact cagaaattcg cttgcctctg cctcccaagt gctgggatta 7980
agggcctaag ctttttcata gccactgtgc ctcaagattt ctatgtcaag atccaggtca 8040
gaaacttgca tcttctaccc tcaagatctg aatcacagat gagacctcca attctggatt 8100
gtagttcatt ccagatatag tcaaattgac atccaagatt agccattaca gtcccagacc 8160
tcacactcta tacctgcaaa tggaatgcca ttccttggtt aagatcatag gttcagcttc 8220
ctttctcctg gtagaaccca ttaagctaga acctgagatt cctagagtcc ttccctatgg 8280
taacgaggca tctcagtacc ttaagccttc agctaatgac ttttgcttac cctggatatt 8340
cctacccctt gacctaaaac tatataaacc ttgaatcacc ccaagttaag ttgatctgtc 8400
tccctgaagc tggtcctggt gtgcccttat tactcactgg gcttctggac acctctccca 8460
ccctccacac cctttctaac catcatctct gctcctggga ggggacaggt gcagggaggg 8520
tcacatttag tcttcttgcc tcaacctttt gaatgctgtc caccgcctgc ctcaccacat 8580
ctgtcatttc ttcttttgca atatatgtag ttgaactcaa aatttccctg ttagtacgac 8640
tttcctgcac acacatttgg attgctaggt ttttttttact ttagtagttt tatcattact 8700
tcgaagccct atgaagatat ttatgttctt ccttggcccc tgggtcatct ctcagccagg 8760

tccctaggta caacctctat attataggac ccaggctccc tgtctggttt cactttctta 8820
acctttgaag ttatcattct aggtaatatt taaaaaaaat agtagtgcac taagaggctg 8880
ggtgtggtgg ctcatacctg caatcccaac cttcaggcta agacaggagg attactgtaa 8940
attcagagtc atctgtggga tacacaggga attccaggtc aatctggtct ccagagcagg 9000
caggcctaca cagagaaact ttgcatatat atatatttta aaagaaagaa tggaaggaag 9060
gtcagaccac attttattag tgagcttggg aatctgcatg ctttgtgact gctggatgag 9120
atgtgatagt gtatcctaaa tggatggatt ttatcatctt ttcttccaga acacaaacag 9180
ctgcttttca tcttcctgct tgcctagctt ctttgggcta gatggttctt tgtagctctc 9240
taccttttaaa ctctacccca ggtaaagttc cgagtagtgg cctatcttta tgttgatcaa 9300
tgtttgatca tccctgaaca gagagaggga atcccctctg tggtttgttt gtttgtttgt 9360
ttgtttgttt gtttgttttt gttttacaag acagagtctc tctgtgtagc actggctgtc 9420
caggaactca ctttgtagac caggctggcc ttgaactcaa agagatccgc ctgcctctgc 9480
ctccccaagc actgggatta aaggcgtgcg ccactaccac caccacccag ctccctctgt 9540
gttgtaaaac agctgctcct gttgctttaa ggctgaggca aggcgctaca ctggggaaga 9600
gaaagaggca aaggtgaata aagcaagata aaactgccat agaatttttc aggccaactt 9660
ttttttgtt tgcctggttg ctatggactt ttatctattt tttttttga gtgcctacaa 9720
agtttgttct gacatctcag cttattttat cagtgtttct atggagagat actatcgtag 9780
agcttccttg tttactgaat tcctgatgtc actcctgaat ggcgtatttt aaaaatcatt 9840
atttactgat ccttcgggac cacaagataa gtgagaactc cagatattca gtctctctgt 9900
ctaaagcaca agaggtgggc aagactaggt tagcacctcg actgtgcacg ttgcatttaa 9960
tgcagatgtc tggaagcaga tcacagagcc gctgcctggg acatgcacgg tggtcagcag 10020
agataatgtt ccctgccttt cacatagacc tccaaactct gaatgctgct ggagaacaga 10080
gaggtcagtt aagtgaactc tcttcacacc ccagggcctg cttgaacagc tttcgttaaa 10140
gaactaccaa gcaaacaccc tggtcccagg gcactgcctt gccccaaccc caaactgccc 10200
cctgacatag tatgaaatgc tgctctgggc tgcagactga ctacactgta gccagagata 10260
attgtatgaa taataaaaac aaattaaaaa gagatattat ggctggctgt acagtttagt 10320
agtgtagcac agacttacca cgtgcaagtc cagggttcaa ccgtcagtac caaaccctcc 10380
ctccacaaac ccggaggtat cttcttatca taccgcagtt gcttggttag taaagcctgg 10440
aaaatataat acataactat aaaagtgtga tgagacactg ttatgattag gtgagtatat 10500
gatgagacac tgtgcttgct ggcatcactg atgtctggcc tgtgagagtt gagactgacc 10560
cttagctaga taccagtaac agattctgat aattgtctga taatcctcat ttagcccaag 10620
gtatgaaatt gctaactgtg cacttcagag tcataaatac atttaaaatc ataggtttca 10680

ggtctgggaa cacatttgaa ggagacacat tccaaaaata aaaaagaaga gggggaggag 10740
agatgagggg gagaaggagg gagaagagga agaggaagga gaaggaggac atgtcttaga 10800
cataaagggg tcaatgggaa cttatttata atcatgaaat cttggcaacc agcccgattt 10860
tcaatgatgg gagagtagac ccctctgctg ctccatggat gataattcca aataaacatg 10920
gtgatcagag attgagggca atggagtttt aaaatcaagg gaaaaacagc aagcgtgcat 10980
gcttgtgtca cttgttcgtg actgaactag tgtgccctgg caaggcctac agggaccccc 11040
acacaggcag tgatgtgaga ccaggtgacc ctccagtgtg actgtgtgtt tctcatgctt 11100
ttgggggctt cagaaaagcc ctcagaccaa ggatctggac ctcatctctc tggagtctgg 11160
tctggggaca gctggacagc cctcgtgaga actgatgtgg agaaggcagg gctcaatgcc 11220
cctcactggg gctcttgggg ttttcatgtg gcagcagtat ctgtagacca ggctggcttt 11280
gaactgcctg cctcttcctc ctgagtgcta agattaaagg tgtgcaccac caatgttcct 11340
attccaggaa tgtcctcaat caatgacttg taagtgtgga tggtgccaca cccttatcca 11400
ctggtgggga tcccctaggg cgggtcccct tatgtcctgg aggtctcagg gacaggtgat 11460
attcttgaat ccatcttgat catcacatct catggttctc agtctgctca gccctctgtc 11520
gtcacctgaa gtcatcttcc aaataaccct gattttcatt tgtcctagga tacttgtctc 11580
agggtctgca tctggagaaa tcacatgaga acatttggga caagacaaga agaggactgg 11640
gtggcatcca gggagcaaca agggaagcag gtgatgttgt gtggcccagg gcccttctcc 11700
tcagcctctc ttgttccctg cctaagcttg ggcggattcc cctctgagcc cacccgagcc 11760
cctgggacac tggtggaact cagtaggagc ccctccctgc agctgtctca acaggtagct 11820
gcatgagtgg ccttgaagca attatcagca attcagccct ggcaatagag gccaaggtcc 11880
tggcctgtct tggtgatagc aagagcccaa ggaaagactg gaagtttcct actggaaaga 11940
agcagaggat gaaccatgta cctgggccca ggttgggtgg gacttgccac tcagagcccc 12000
taaccagggt tgttcagagg actaggccag ggccaggacc aagaaaggga tagaacgggc 12060
atgaggagga agggtgaagg gatccaagga atctctggtc ctgttccctg ttaggacatt 12120
tgtcatggaa tcactctcgc ttagtgtctc tgttatctgg gtgctaatag caactattca 12180
gttgctagga tgttaggtga gtctgaacct acccttgatg ttgatctgaa gaggcgatgc 12240
gttagactgc aggttggagg ccaagtccag gacagtgttg atattctgga tctccaagaa 12300
gcctccaagg ccaaagccag gccagtgtct ggtctcgcag aggaacagct ctgcatctct 12360
tgcccggttg gctctaacta ccacattaga cttcagttgc gtcaaaaaac gaggggaccc 12420
cagcgccttc actaggaagt tgacctcaga aggaggagat ggaatggcac catctgatgt 12480
aagggaagag aaaataaatt attaaccagt acggcccagt cctattggcc ccatgacaga 12540

cgagggttat cactaagagg aggaagctgc cttaatgtgc aaactcaggg gccagtcctc 12600
agcttccccg gctgtctcca aggcctggtc ctgcttttcc ttgatcactt cctggctctg 12660
ggatggcagc tgcctggcag ggatggctgc tctgggccct gctcctgaat tcggtgagtc 12720
tgttgcttgt ggctactcct tggtgcctcc cattagggca aagtgatacc tgatacctga 12780
tactgggtac ctgataacgg ggaagggtcc cacggctgtg ggagggttcc tatgcccaaa 12840
gataagtgct ggtggagggg tctccaggtc aaggggttga agggatagag gtcagagagg 12900
caaagggatg gggcctttgt ctgaggttaa atggggacca agtcaggtgc tagaggtgga 12960
tcccagtgaa cagcgcctga aatattctgg gcttgggagg aggttttgct accatccttg 13020
tttgctctca ggcgatagca ttggccaatg caggatgtag gagtgggggg ctcttataca 13080
gactcttgta caaggaaccc tgacctcggg gtagagctca gcctggagac tcaaactgac 13140
agcaataaag gtcgctatct cctactctcc cctgcagcac gacccтttaa agccacactc 13200
tattggatca cttccttttc tgaatagccc cctcactgtc cattggggga gtgcccctcc 13260
attggcaccc taagcatagc acgagccccc acaagcctcc cgcagcactc ccagcccctt 13320
actgctggcc ttcttaccca tagactccct agcctctcac tctccagaca gtccctggct 13380
gtgccaacca gccttagggc ttatggatgc tatcggttct ttctgcaggc ccagggtgag 13440
ctctacacac ccactcacaa agctggcttc tgcacctttt atgaagagtg tgggaagaac 13500
ccagagcttt ctggaggcct cacatcacta tccaatatct cctgcttgtc taatacccca 13560
gcccgccatg tcacaggtga ccacctggct cttctccagc gcgtctgtcc ccgcctatac 13620
aatggcccca atgacaccta tgcctgttgc tctaccaagc agctggtgtc attagacagt 13680
agcctgtcta tcaccaaggc cctccttaca cgctgcccgg catgctctga aaattttgtg 13740
agcatacact gtcataatac ctgcagccct gaccagagcc tcttcatcaa tgttactcgc 13800
gtggttcagc gggaccctgg acagcttcct gctgtggtgg cctatgaggc cttttatcaa 13860
cgcagttttg cagagaaggc ctatgagtcc tgtagccggg tgcgcatccc tgcagctgcc 13920
tcgctggctg tgggcagcat gtgtggagtg tatggctctg ccctctgcaa tgctcagcgc 13980
tggctcaact tccaaggaga cacagggaat ggcctggctc cgctggacat caccttccac 14040
ctcttggagc ctggccaggc cctggcagat gggatgaagc cactggatgg gaagatcaca 14100
ccctgcaatg agtcccaggg tgaagactcg gcagcctgtt cctgccagga ctgtgcagca 14160
tcctgccctg tcatccctcc gcccccggcc ctgcgccctt ctttctacat gggtcgaatg 14220
ccaggctggc tggctctcat catcatcttc actgctgtct ttgtattgct ctctgttgtc 14280
cttgtgtatc tccgagtggc ttccaacagg aacaagaaca agacagcagg ctcccaggaa 14340
gcccccaacc tccctcgtaa gcgcagattc tcacctcaca ctgtccttgg ccggttcttc 14400
gagagctggg gaacaagggt ggcctcatgg ccactcactg tcttggcact gtccttcata 14460

gttgtgatag ccttgtcagt aggcctgacc tttatagaac tcaccacaga ccctgtggaa 14520
ctgtggtcgg cccctaaaag ccaagcccgg aaagaaaagg ctttccatga cgagcatttt 14580
ggccccttct tccgaaccaa ccagattttt gtgacagcta agaacaggtc cagctacaag 14640
tacgactccc tgctgctagg gcccaagaac ttcagtggga tcctatccct ggacttgctg 14700
caggagctgt tggagctaca ggagagactt cgacacctgc aagtgtggtc ccatgaggca 14760
cagcgcaaca tctccctcca ggacatctgc tatgctcccc tcaaaccgca taacaccagc 14820
ctcactgact gctgtgtcaa cagcctcctt caatacttcc agaacaacca cacactcctg 14880
ctgctcacag ccaaccagac tctgaatggc cagacctccc tggtggactg gaaggaccat 14940
ttcctctact gtgccaagtg agtagatctg aggggaacag gtgagagctg ctatgccccc 15000
aggaaccagg ccagaaccta gctccaccct tgggagccag ggacagctcg tatgtgcaca 15060
tatcagggcc atggcctgtc caagtctatt taagtccctt cttggagctc actcccatct 15120
tattcctgca ggaattttgt cctaccagtc tttccagctc caatccatat gatctttcca 15180
tccatgatgc tcctggtatc aacttaataa tttttagaat tactttaact tcacatgaat 15240
gaatattttg cttgtgcata tgtatacgca ctgcttatat atgtgcctgg tgctgaagaa 15300
gccggaagaa gttgctagat tttcaggaac tggagttgag gtcagtcata gcggccaggt 15360
gggtcctggg aaccgggctc tggccctttg cagaaatacc atgaaacgtc tgcgtcctct 15420
ctcctgccct catgtagtct tagtttaaat ctcaaagcga tgtctcaggt agtgtttgtg 15480
tttgtgatgc ttccttttct ggactctgtg tcttggtctg tagggacttt ggtagcctca 15540
caactggcta gaaatatgtt catctgggct taggtggaac tgtggttagt ctccagtccc 15600
aggcatcagc acagtttttt ctacaacctt atgctgttga ggttctgctt ctggctttgt 15660
ccattttggc tggcacaaac aggatgccag tgagctcatc agacagaggg aaggttggtg 15720
agagggccag aggtagagga ggctcctgga gaacatcatg gagagtgaag tgcctcaaat 15780
ggccttgtcc actctagagc aggcgagggt tacagcaggt aaccacagct gagtgttctt 15840
atgaaaacag ttttgaccct gcaagcccca gacttcatag tctttagagc catcagatga 15900
gagcagaaag cttttgctgg ctctcattgc tactggctgt ctatccccgt ttgagtctcc 15960
agtgcaagct acttcctaga gtatccatgc tgtcccctag atcggacagc agagaagggc 16020
tgtggagagg catcggggat cagccacgca aaagacaatt taaaaaatat tatttatttt 16080
tatttataca ggtacactgt acctgtcttc agacacacca ggagagtaca tcagatccca 16140
ttacagatgg ttgtgagcca ccgtgtggtt gctgggaatt gaactcaaga cctctagaag 16200
agcagtcagt gctcttaacc tctgagccat ctctccagcc ttacagtgta gcttttgttt 16260
ttgttgcttt tatgttgttt ttagacagag tcacactatt tgagacggtc ctcaaattca 16320

cgcccttgcc tcagcctcct gagggctgtg gctgcaagcc taagccatca tacttggctt 16380
gtactacctt tattttgatt ttgaatgctc ccgactcctg gtgagtcagt tatgttaatt 16440
ctatagactg gaaacctgag gctcagagtg gtatggtaag acgggcaagg ccacacagga 16500
atccagctct ctttgacggc tctgttgatc aatatactca cttgttcaga cctcagaatg 16560
tgtataaaga gcttggtgtt ggtagtctat tcagtctcca cagaggtgtg ctctgtagaa 16620
agggtttagt tgaagggcac aggcccctgt ccaagggcat tctctgggtc tgtgagctcc 16680
agggctcaac tcaacattta ggggtgattc tagctctggg aggggaaagt gaagaacagc 16740
attgagatct gtgagggaga tgggcatggc tcagttctgg gctcatcact taatggtgat 16800
gctcatttga caggtctgga aggtttggct atgtgagggg gcataggaag catcacctgc 16860
ccaagggaac cacattcagt ggactagggg accatatgag actaccttgt gaggagatag 16920
tcattttgaa ctctctgggc ctggtattgt ggagacactg ctcctccaat agcggggaga 16980
ggagctgggg cagggagggg ccaaagagtc caggcagggc caggaaaggt tctttccctt 17040
tgtggtttcc ccctagtgcc cctctcacgt acaaagatgg cacagccctg gccctgagct 17100
gcatagctga ctacggggca cctgtcttcc ccttccttgc tgttgggggc taccaaggta 17160
agtgaggtag ctgggggggc tactgaaggg ataattttgg cacagagata ataggtagga 17220
ggagggagaa gccatggtga gtgtatccag gatctggggg cctggcataa gggggctgca 17280
ggcaatgctt cctacctcac tgctctcatc tctcaatgct acccaggagc tctggttttg 17340
tgcctttggc tgggaaaggg aaatgaagca tgggataagg ctgttattgg agtgaggaag 17400
caatagaagg acaggaatgg gagaaggtta caccctgagg ggaggagggg aaaagggttc 17460
aacaggaggg aggctcaggg tttctcttcc cagggacgga ctactcggag gcagaagccc 17520
tgatcataac cttctctatc aataactacc ccgctgatga tccccgcatg gcccacgcca 17580
agctctggga ggaggctttc ttgaaggaaa tgcaatcctt ccagagaagc acagctgaca 17640
agttccagat tgcgttctca gctgaggtag gggccctgca gagtccctgg ttctatgctt 17700
gcaatcccta atggtgtggg tctattccag tcaaatctac aaactggctc tacttgttcc 17760
tgactggccc cgggcagtga acacctgtgc ctagctgtgg cgcttgtgtt agaggctcct 17820
gcagttcatt cctagagtgt gtggccactc agtatgtggt ccgtgagctg gctgtgtgct 17880
tgcagcgttc tctggaggac gagatcaatc gcactaccat ccaggacctg cctgtctttg 17940
ccatcagcta ccttatcgtc ttcctgtaca tctccctggc cctgggcagc tactccagat 18000
ggagccgagt tgcggtgaga gcaagaggga cacagtgaga gtgactcaga gcctaggaca 18060
cctccagaag gcttttcaaa gcttcccgag tgtgggcaca ttaaaatagc aagttggaca 18120
catccagatg gaatcccttg aagggtagcg tttcttgggt gtgttctatg ttgaaaggct 18180
ttcttcctgc tctctaatat attccaactg tctacatgca aagctaccat ttaaaaggcc 18240

atgcaatgca gttctgggaa ggtgcagcca ggtacccctg cattctttgg ttccatgggc 18300
ttgcccctga gagcatggtt tagcatagag acttagatgt gggttcttca ttgaggtggg 18360
tggtgtgtga gcaccaatga tgcctgccca ctcctcagcc accctggaga gtacaaaggg 18420
tctgggcagg tgtcctggta gccagccctc ctcactgaat tgcaggtgga ttccaaggct 18480
actctgggcc taggtggggt ggctgttgtg ctgggagcag tcgtggctgc catgggcttc 18540
tactcctacc tgggtgtccc ctcctctctg gtcatcattc aagtggtacc tttcctggtg 18600
ctggctgtgg gagctgacaa catcttcatc tttgttcttg agtaccaggt aagaagggag 18660
gggttcttca tactcaacat cctcattaga caaagttctg cacagactca ctggaattct 18720
ggtcaattta tacgtgtagg aaatagcctg ggttggcaca aattcattca cactcattga 18780
gccatcttga acttgcttcc agttaaaccc atacagcatc cagtaagctt tgtaatggat 18840
tagaggtacc tctttcctgc ctttacatta ccaggggcgg cattccatgg tataggcaca 18900
agccagagtc cagatagtct ctctttgctg tcaaacactt ggcgtgacat gaacacttgg 18960
tcgtttccac atctagaacg caccagtggt tctttacatc ccaacataga agcagagagc 19020
gtggctgtga gctgttagta ggctcttctg tccacggaag gtctggaagt tcctcagatt 19080
tggccaggaa tccaaaccct aaccacccca atgctgacct ctaaagtttg gtgaccttgg 19140
gctggagaaa tggctcagca gttaagagca ctgactgctc ttccaaaggt tctgagttca 19200
attcccagta actacatggt ggctcacaac catctgtaaa gggatgtgat gccctcttct 19260
ggtgtatatg gaaacagcta cagtttactc atatacataa agtttggtga ccttggcaca 19320
cccgtgtact ctgtctcttt gcccatgcag aggctgccta ggatgcccgg ggagcagcga 19380
gaggctcaca ttggccgcac cctgggtagt gtggcccccа gcatgctgct gtgcagcctc 19440
tctgaggcca tctgcttctt tctaggtgag caagggctgt ccttctccac ccgggatggg 19500
atttgctagg ttattctaag agggagccca ggctttcaga aggcagtggg tgttccctgc 19560
tttaagctgt ctgtgctggc atgtggccca tgatgccaga atgcccgaca gaccctgtgc 19620
cctcgacagg ggccctgacc tccatgccag ctgtgaggac ctttgccttg acctctggct 19680
tagcaatcat ctttgacttc ctgctccaga tgacagcctt tgtggccctg ctctccctgg 19740
atagcaagag gcaggaggta agttcaactg ggccaggaca agggacttac cctgccagtg 19800
tccctatatt ctctggaaga tgtggcacag aggtagccag aagagtttga tgggaggcag 19860
ggacagtatt ctgagagaga atgtttgggg ctctgtgctc accaatttcc tgtaaaaaga 19920
gaatttcttt ttagttattt gtggtaacat catcaacgcc cctaaaagta tgtaaagttt 19980
acaaaataaa ttgtaaataa aaagttaaca taaatttttt gatgacggaa aattcagtat 20040
ttgattaaga caggaagtaa gctgggtgtg gtggcccatg cctttaatcc cagcacttgg 20100

gaagcagagg caggcggatc tctgagttcg aggccagcct ggcctataaa gtgagttaca 20160
ggacagccaa ggctacacag aggaaccctg tcttgaaaca aacaaacaaa caaacaaaca 20220
aacaaacaaa aaccaaaaag acaggaagta aaagcaacaa aaaactgcgt gggggctgta 20280
gagacggctc agtggttaag agcactggct gttcttccag agatcccgag ttcaattccc 20340
agcaactaca tggtggctca ccatccatac tgggatctga tgccctcttc tggcagcagg 20400
tgtccataca gttagaccac tcatacaaaa tcttctgggc ttccttgaat gaggtggatc 20460
ctgtagtctg ccttggaccc agtcttgagg gcctgtcatt ctctaggcct ctcgccccga 20520
cgtcgtgtgc tgcttttcaa gccgaaatct gcccccaccg aaacaaaaag aaggcctctt 20580
actttgcttc ttccgcaaga tatacactcc cttcctgctg cacagattca tccgccctgt 20640
tgtggtacgt gggctgaagg gctgttccac ttttgtacca ctttgggagg gaaaccgggc 20700
agagcatggt ggcatgggag gctgcccagg cccggagcag acacttggag ctagagcttg 20760
agcctgtcca actctaggac gtttcccagg atgcccaaca aagccattca aatttgaggg 20820
aagatgaagg ctgtttgggg agaggttctc acgtgccagt ttttccctca gctgctgctc 20880
tttctggtcc tgtttggagc aaacctctac ttaatgtgca acatcagcgt ggggctggac 20940
caggatctgg ctctgcccaa ggtgagcctg gccttttctc agcccttttgt cctgggaggg 21000
gcagcagtgc ccaataggtg gagcggtggt ggtggtggtg gtggtggtgg tggtggagct 21060
tgagaggggg acatagcaca aggcttagcc ccatgcagag ttgctctaag tggaccgtga 21120
gagagaaagc acatccatgt tgtaagtgtg agcgctgagt gctggctcag ggtcacagta 21180
gatgtcctgt gctggaggcc tatccacatg gccattcaca cagggtgggg cgccacttcc 21240
ttctatgtca gttcctcacc aatagctggt ttcggattta ttactttatc tgtacgagtg 21300
ttttgtctgc gtgtatgttt ttgtgccatg tgggtgcctg gtgcctgcct gcagaagtca 21360
aaaggagggt gtcagatccc ccgggactgg aattacagat ggctgtgagc caccctgtgg 21420
gtgctgggaa ctgaacccgg gcattctgcc gagccaactc tccaacctca gcacttgtta 21480
tttttctgtg ttttttttt ttttttttt ttttttttgt gtaggggaat caaatctggg 21540
atctcccatt tgtcttgttt cgatctcttg agagtcctag caacaccgct gtctggcttt 21600
atagtttcga tttgcatttt ctttctttt ctttttaaag atttatttat ttattatatg 21660
taagtacact gtagctgatt tcagacaccc cagaagaggg catcagatct cattacaggt 21720
ggttgtgagc caccatgtgg atgctggaat ttgaactcag aacctttgga agagcagcca 21780
gtgctcttaa ctgctgagcc atctctccag gcccctcaat ttacattttc aacaattaga 21840
aatgttacat acctttttcat gtacatgttg atcactatat atcttattta agaagaaatg 21900
tgctgacttt gctcggtttt tgaattggct ttttgttgtt gctgagcctt ggagagttcc 21960
ctgtggattc tggaggttgg tgtcttctca gagacctgat tatcaaatgg ttttcttttt 22020

ctgtgggctg ctctgttatt ctagtggtgg tgtgccttgg tatgccaaat atttaagcat 22080
atccatggat tctttttctc ttttattgtc tgaatttgat ggcatattaa agacataatt 22140
gataaacaga aagttattaa gtttgtctgg tttctattaa ggtttttttat gactttagaa 22200
cttctgttta agtctttgat tcatttgaga tttgctcatt tgtttttga aatagggttt 22260
gtctgtgtag ccctagctgc tctggagctc actctgtaga tcaggctggc cttgagttca 22320
gagatccacc tgcctctgcc tccaagtgct gggattatag gtgtgtgcca ctaccccact 22380
ttgaattgac ctttatatat gatgttatga aagtggacaa attttaattc catccagctt 22440
tcccaggact gtactaagaa gtacagctct cctccatccg atggtttggc agccctgcca 22500
gaggtcattc aagcatgtct gtgcatgact ctttattctg ctccattgaa atttcatgcc 22560
ggcttccgtg tcagcagggc cctgctttga ttcatacgga gttgcaaacc agaaaatgtg 22620
agacttgcaa atttgttctt tgtcgatttc tcttgggcta tttgagttct tgtgagatta 22680
cacttgaatt ttagcttgac atttttagat tccttcaaaa accatccttg gcatttatgc 22740
agggattgca ctgaatctgc agatggcttt gccttgatag tactaatacc gtcacaatat 22800
ttgtcatcca gcccatggac acacgatgta tttttttttc atttttttct ttaatttctt 22860
ctaagaacca catctccaaa tttttaattt tttttttttg agacagggtc tcactacgta 22920
gccctgcctc actgtgtact cacatgtaga tcagactggc cttgaaccca cagacatcgc 22980
ctggctctgc ctctcaagtc ctgggaccaa aggtgtgtgc caccacacta ggtctgagcc 23040
actggctttc cacatgctaa gcatgcactc ttaccactca gatgcaccct gagccctcct 23100
ctctgaagga tagttttgct gtaaataagg ttttttcttc cctttagtac tttgaataca 23160
tgaaccgcag tctccaacgg cagatgggaa aggtggaagc agctgcgtta gtcctttgtg 23220
acaagccatt ttttgtgtgt tgtccccaga gctctctgag gttggctttt gacagctgta 23280
ctacagcctg ccttggtcaa ggtttgagct tgtccttta gacgtcccag gaattccttt 23340
aaggttgata ttcgtgcctc tctttcatcg atttgggga gttttggcta ctgcttcttc 23400
aaagatcaca tccagattct ttcatttttc ttcctttct gattttgttt ttgtttttgt 23460
tttttgtttt tcgagacagg gtttctctgt atagccctgg ctggaaactc actttgtaga 23520
ccaggctggc cttgaactca gaaatctgcc tgcctctgcc tcccgagtgc tggaattaaa 23580
ggcgtgtgcc accaggcctg gctttttttct gaaattctta cagtgcatct gtgggcctat 23640
tggcatcgcc tggggccctc tctgtgtgta tcttggtgtg ccctttatg tttgtatctg 23700
ggcatgttcc tcgaagtaca tagtgctgtg tgtacctcag cctgtatctt gtatatatgt 23760
acacatgtag tctgtacatt tctgagtagg tttctgagca tgtgtctctg agtgttctga 23820
gcatgcatct ctgagtgttc tgagcatgtg tctctgagtg ttctgagcac gtgtatctga 23880

gtgcgtccct gagtctgcct ccgagcatct catctacctc gtgtacctct gagtgtgtct 23940
tctgcttcca gatacatctg catggacttc tgagactgtg ctctgacctg gggcggaccc 24000
actgtggata tttcctacac tcacagatcc tcttctttcc caggattcct acctgataga 24060
ctacttcctc tttctgaacc ggtacttgga agtggggcct ccagtgtact ttgacaccac 24120
ctcaggctac aacttttcca ccgaggcagg catgaacgcc atttgctcta gtgcaggctg 24180
tgagagcttc tccctaaccc agaaaatcca gtatgccagt gaattcccta atcagtaagt 24240
ggttggtctc cccgacaccc tggcttgttc cttctctgct ttctctctcc attcctcttc 24300
tctcttcctg catgctctgt ttctgcagct aacaaagcca ggggaggctc cagtgcaagg 24360
gtaaggaagg agtccccagc agactcattg gctccacctc ctcctctcca ctgtctggcc 24420
tcaggtctta tgtggctatt gctgcatcct cctgggtaga tgacttcatc gactggctga 24480
ccccatcctc ctcctgctgc cgcatttata cccgtggccc ccataaagat gagttctgtc 24540
cctcaacgga tagtaagttt ggggctacag gaggctcact gcccattaca gcttagggaa 24600
actgaggcag gagaaaagaa aggctctcag tctcccatca aacccatagg gtccaggtgg 24660
tttaggggtt aggcactcac actatcagtg tcccctggag tattacacct ttgtttgcag 24720
aacatgttgg ttgtgggcag tgggctatgg agttggaagt ggagctatgg ccctgcatat 24780
ggagctgctg tgtttaacaa gtgtgggaga tcccatttct tgaccccaca actgggggtg 24840
gcaggtgtaa acctcttaga actggggact ttagatttgg gcacagaatg ggagtcagga 24900
caggagctgc cttgcctggt gtgtcactgc ccagagtcct ccctctctgc agcttccttc 24960
aactgtctca aaaactgcat gaaccgcact ctgggtcccg tgagacccac aacagaacag 25020
tttcataagt acctgccctg gttcctgaat gatacgccca acatcagatg tcctaaaggg 25080
taggttccga gggtggctct tgctggagac tggggagact agtgggttct agaaatggta 25140
gacacagagg aggcaagagt gcctagccaa gccctttctg gggcacagtg agtggactga 25200
caggacaagg tctcgttccc tctaagcctc tactctgtcc tccactttgc aggggcctag 25260
cagcgtatag aacctctgtg aatttgagct cagatggcca gattataggt aagtgtgata 25320
tggtttgggg aggagatctc aagtcagtca gctgttttag agtcctctaa gagcacccat 25380
gcatgtggct gacgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgta 25440
agttagcagg gtgtgagggt aggtgtataa ctgtcctggg cctgtaagca ccggttctcc 25500
ttctcaactc cacgagctaa tacttcaccc attctttcac cccaagtcca ttgccactgt 25560
gaaatgtgtg gcagcctttg taacagctga cgtcattatt gagagccacc cacttcaaca 25620
catattcact ctcgtgtgta tcatatgtcc atgtcagcag caacttctgg ctaaatgaag 25680
taggatgcct ttgtgtagtg gaatctcaag aggcatcaac agtaactagg gagtgactct 25740
gacaggtggg gaggagctta atccagcagg cagcagaacc aaccaatcat ctgcatggag 25800

ggagccagct gttgttagat ctgtgcacac cactatgagt gcaggagctg gcaagacggg 25860
tctgggtgct ttggatgaat tcagtttttt acctaaataa ctccaagttg gaatcactag 25920
ctgtaaactg atggcagcac atctagtcca ccctctaagc taactcttct atgaagttcc 25980
tatcttccaa gaagtcacac caccttaccc taggcaacag aggcatgtgg tcaaaaggga 26040
tcggggtatg gcagacagag gaatggattg tttcttggag ggcaggtctg catctgtcat 26100
ggccctaagt atcccacagc agcgcttttc tttctttttt aatttttaat ttttttggtt 26160
tttcgagaca gggtttctct gtgtagccct ggctgtcctg gaactcactc tgtagaccag 26220
gctggcctcg aactcagaaa tccatcagcc tctgcctccc aagtgctggg attaaaggca 26280
tgcgccacca ctgcccggca gcagcacttt tcaaaatgag agttcccctc tctcctctaa 26340
cagtatcagc attatcagga gatggctagc catgcccacc ccttgcctag ccatgcccac 26400
cccttgcctg agccatgcca gaccatatct cggcacgagg atgaccatcc ttctgggtgt 26460
gagcaactag taccttaaga ttacgcattg gcgactattt ctcatctgtc tttgttttcc 26520
ctgtgtgtgc ctagccctcc tttctttagc ggatgcaaca ctgctgaaca caattcacag 26580
ttgtttattg tacttacagg caggatccca aagtcacagc tttaataatt cagtcatttt 26640
tgtttgcctg tcttcctttc tagtctgttc ccacaggaat aggcaaactg aaaattaatt 26700
tttttgagac agagtctcat gtagcccaca ctagtctaga actttctatg tagtgctgaa 26760
ctcctgaatc tccagcctcc ccaagtactg ggaccacaaa tatgaaacaa cacactccaa 26820
caagaaagta ggctcaattt ttgattaaaa tccagtgcat attctgaaag cacacataga 26880
aggatttggt ttcaccaggg agatatttta gtactttagt ttagtttttt ttttttcttt 26940
tttctttttt tgagtttctc tgtgtagccc tggctgtcct ggaactcact ctgcctcgaa 27000
ttcagaaatc cgcctgcctc tgcctcccaa gtgctggcat taaaggcgtg cgtcaccact 27060
gcccggagag attttagttt ttgctttgct tttgaaacag tatcttactc tctagcccaa 27120
gctggccttg aatttgaagt aatttcccta cctgtctgct ggctgctgga atttcaggtg 27180
gtacataagg catgatcatt caagcttcat ctatcaaagt caggccttgg tctcagtggc 27240
agaagtagag tttgaggatg ttgatgcaaa atgtaaatgc aatccttgag gctggagaga 27300
tggctcagag gttccagagg acccagctct gactcctgca ctggtgcagc agatcatgac 27360
tgcctttaag tccagttaca ggcgatccca tgccctcttc tggcctctat gggcaccagg 27420
aatgcatgtg gtacacagac atacatgcag gcaaaacact catacatgtt aaaaataaat 27480
aaggaaatgc aggtctcttg tagagaatcg atcaggaatt tcaagtgctg ctggtggcac 27540
cctgaaccca atgagatcac atctgtgtta atttctccgg gatacgccaa tgggcactga 27600
tctttgactt tcagcctccc agttcatggc ctaccacaag cccttacgga actcacagga 27660

ctttacagaa gctctccggg catcccggtt gctagcagcc aacatcacag ctgaactacg 27720
gaaggtgcct gggacagatc ccaactttga ggtcttccct tacacgtgag aactagaggg 27780
ctgagtcagg ggtgtgggag gaaacagcca cacaagagat gttgggggga ctgtaggctt 27840
tgagtgatcc tgtgggatga ggaccaactt tgtctcagct gtctcctggg gtagcttgcg 27900
gcctgtccat ttcttacaca ggtcagcacc caactcaaat ctgggtccct tgttttcctt 27960
gggagatggg atgctcatat cacctgaagg gaggattgat gctttgacca caccctgatc 28020
tttggcagga tctccaatgt gttctaccag caatacctga cggttctccc tgagggaatc 28080
ttcactcttg ctctctgctt cgtgcccacc tttgtggtct gctacctcct actgggcctg 28140
gacatacgct caggcatcct caacctgctc tccatcatta tgatcctcgt ggacaccatc 28200
ggcctcatgg ctgtgtgggg tatcagctac aatgctgtgt ccctcatcaa ccttgtcacg 28260
gtaacccaca gagcgggcct tggaagttga cgatctacac tcataagcta ccctcattct 28320
atgtagaatc tagacatccg gctggcttgc tttcttgtac ccacacccca ccttctccct 28380
tgtttcctga agttcatttc tcttgcctgt aggcagtggg catgtctgtg gagttcgtgt 28440
cccacattac ccggtccttt gctgtaagca ccaagcctac ccggctggag agagccaaag 28500
atgctactat cttcatgggc agtgcggtga gtggggaggg atggcctcac cctgcgatcc 28560
acctgagcct ttatgtcctc ctgtgctgac tcctggctgt gactcctgcc aggtgtttgc 28620
tggagtggcc atgaccaact tcccgggcat cctcatcctg ggctttgctc aggcccagct 28680
tatccagatt ttcttcttcc gcctcaacct cctgatcacc ttgctgggtc tgctacacgg 28740
cctggtcttc ctgcccgttg tcctcagcta tctgggtgag tacctgtgca cacccggcca 28800
agatgtcaca actgtgagca ttgatcaaaa tggtgcctgc tcccctggaa aacttagaga 28860
tttcaggctg agggttttac catacatcct actttgggag cttttgtttt acatttaata 28920
tgcagcaagc atctttctct gtgagttgat tgtgtcttaa acactgtgtg ggctgctgaa 28980
cttttctgat agacgtttat gcacatacaa acacacacac aaatggacac acatgcacat 29040
aaacacacat actcacatag acacacactc agacacacat gtacacattc acacagacac 29100
acatagacac acatatatac agagacacac agacacaagc atacactcat acacacagac 29160
acacatatag acactcacag acacactcac acaagcacag atacaaatac acacacacac 29220
tcacaatttg aatgcacaga cacacagata cagatacata cactcatact cacatataaa 29280
cactcacaca cagacacaca aaaacatgta caggggctgg agagatggct cagtggttaa 29340
aagcagtagc tgcttgctct tccagaggtc ctgagttcaa ttcccagcaa tcaaatggtg 29400
gctcacaact atctataatg gtaaccaatg ccctcttcag gtgtgtctaa agacagctac 29460
agtgtactta tataaatgaa ataaataaat ctttaaacac acacatacaa agacacatac 29520
tcatatacac aaacacacac acacacatac agacacacgt tcacagaagc acagacacac 29580

actcacagat acacacagac atgtacataa aagacacaca catacacaga cacacacaca 29640
cacacacaca cacacacaca cacacttctc tgtagatgga acccagaata ttgcacatgc 29700
taggcaagta ctgtaccact gagtcacacc ttagcacaga atacatattt tacaatgaga 29760
attgatgagt gaggtccatg agattgtctg agtagattct gagcctcttg ctcatatagt 29820
aagagaaggt gtattttggc caatcacagt atatatgttt ggaatctcag ggtcctcatg 29880
gcatcaactg tagtcactga ggctctgttt tggaccaatt acagctttcc tcttggcatc 29940
aatttcatcc tggattgtgc tttaggccaa tcacagtctt gcctcttagc ctcacaacag 30000
tctccagcta catcaccagg ggtggtggtg gtggtattca acctacagct gatgaaggcc 30060
taagggccgt ctggctgata gttcttcagg gcagacagaa cagcaaggcc gagtcccaca 30120
ggtggtgttt aggaaagcag attgccccat cctgcagcat cttagcttgc ttacagggac 30180
atgggcatca ggagcgctta caatataatc taaagagatt attaggacga ataatgatct 30240
taatatattg ttaatggtgc tgcacatgct taactcacca agtaccccag gaagaagtgt 30300
ttcccttatc tgcatcactt cctccactct ctatttaagc agcctacaac ttctggtcat 30360
tagactattt ctgatgctat actactgttg ctagcactac agtaactgac cttgttgctg 30420
aatctttgac cttgtcatcc agtatttttct tagagtaaac ctggagatgt tgattattga 30480
tgtcggttgt taatactccc cacaaagttc aatggtgatg ataatggtgg tggtggtggt 30540
ggcattagag tcacctacag gaactcactg actatctttg tggagaagaa tgtgtatgtt 30600
ggggacagtg agggaacagc cctgggagat gttgccagcc cagagcctca gagacacagg 30660
ctggaagttt ctagacctat atggggtgga gagtactgag gactgcaagc ctcccatccc 30720
cagtgatgaa gctgtagtca agaataccct gaagctaggg ctatgcaagc agagtcccga 30780
agggcatgtg gtgagtatag agccctactc cctggttgcc ttgtgccttg ggtttgtagt 30840
tacaatataa ggtatgcttt ttagaggaga gttgaccatg gtgtcagtac cagattgtcc 30900
caggaacaaa gggagagaga gggtgtcagg gatcctgttg agagagcatg ccagctgagg 30960
caggctggtg agggtggtgg aaataccagc tgaaacagct gttggtaatg tgaggcaagc 31020
gcaagcaggt agagccgggc tctatctaga tttgcatacc gctgtgatac ctgcgtcatc 31080
tgtatgcctc tcagaccata gatgtatgtt ctttctcttc cacagggcca gatgttaacc 31140
aagctctggt actggaggag aaactagcca ctgaggcagc catggtctca gagccttctt 31200
gcccacagta ccccttcccg gctgatgcaa acaccagtga ctatgttaac tacggcttta 31260
atccagaatt tatccctgaa attaatgctg ctagcagctc tctgcccaaa agtgaccaaa 31320
agttctaatg gagtaggagc ttgtccaggc tccatggttc ttgctgataa ggggccacga 31380
gggtcttccc tctggttgtt tccaaggcct ggggaaagtt gttccagaaa aaaattgctg 31440

gcattcttgt cctgaggcag ccagcactgg ccactttgtt gtcataggtc cccgaggcca 31500
tgatcagatt acctcctctg taaagagaat atcttgagta ttgtatggga tgtatcacat 31560
gtcaattaaa aaggccatgg cctatggctt aggcaggaaa tagggtgtgg aacatccagg 31620
agaagaaagg attctgggat aaaggacact tgggaacgtg tggcagtggt acctgagcac 31680
aggtaattag ccatgtggcg aaatgtagat taatataaat gcatatctaa gttatgattc 31740
tagtctagct atatggccaa ggtatttata aatatatttc gagtctgagt cttatttctg 31800
ggagcatggg gctgggtggg aagaacaggg cccaacaatc ctccttcttg cccagggtct 31860
tgtagttgcc gggaacatgt ttgtatctct cacccagcat ttcctcccct tatcaaaact 31920
atttccaggg ctggagcact tgttcttaga gagaacatgg gttcagttct cagtggttca 31980
caatcatcta caattccaat gtcaggaaat ttgacacctt ctgatgttca cagacaccag 32040
gcatgtggtg cacatatgta caggcaagac actgatacac acaaaacaaa caaatacatc 32100
taaaaatgat ttaaagaaaa catctttagg gccagtgaga tggctcagtg gttaaaaggt 32160
gattggcatc aatcttgagt ttgacccccc ggaactcata tgatgggagg agggaaccaa 32220
ctcttggaag ggctcctctt acctctacat ccatgcattg gcacccctag cccccagaag 32280
gtaacaacat attaataaag tctccgttct aggatggggc tgtagctcag tgctggagca 32340
gcagcatggc agcctcatgt acatgcagtg tctgtcacct gccatcctca gtactgaaaa 32400
ggacagagag caagagcccc gaccttgtcc ctagatgtta ccacttccag tgacaataac 32460
tgcctttgtt taccactgtc cctgagtaca tttaaaaaaa aaccctccat tccatatcag 32520
catgactgtt aaatgactgt taatatttac ctatagccct aggacagagt gtgacccacc 32580
ctgggctgta atgttttaga agagcaggga aggcaaaggg gacctaatgt cttcctggct 32640
tgaggaggtc acagtacgct gggagtggtt gacctcatct ggaaaatggc attcagtttg 32700
gcctccagtt tcctcagcta cagagcatgt tgcaggcgct gtgtgtctgt ctgaaggcag 32760
acagctctgg gctgggcagg ttttctggca tgggtcttat ggctggagca caacctgaat 32820
ctggtgcctt ggttgcaaca gagacagaga agaagatacc ttgtttgtga agcacagact 32880
ttgttgaata gtgtcgtaga gagtgtttta ctgctgtgag cagacaccat agccaagaca 32940
actcctagtg tcttagagag ggttttactg ctgtgagcag acaccatagc caaggcaact 33000
cctagtgtct tagagagggt tttactgctg tgagcagaca ccatagccaa ggcaactcct 33060
agtgtcttag agagggtttt actgctgtga gcagacacca tagccaaggc aactcctagt 33120
gtcatagaga gggttttact gctgtgagta gacaccatga ccaaggcaac tcctagtgtt 33180
gtagagagag ttttactgct gtgagcagac accatgacca aggcaactcc tagtgtctta 33240
gagagggttt tactgctgtg agcagacacc atgaccaagg caactcctag tgtcttagag 33300
agggttttac tgctgtgagc agacaccatg accaaggcaa ctcctagtgt cttagagagg 33360

gttttactgc tgtgagcaga caccatgacc aaggcaactc ctttaaggac aacatctaat 33420
tgggtttggc ttacaggttc agaagttcag tccattatca tcaaggtggg aacatggcaa 33480
aatccaggca ggcatggtgc aggaggagct gagggttcta catcttcatc tgaaggttgc 33540
tagaagactg gcttccaggc agctagaatg agggtcttag gctcacatct acagtgacac 33600
acctactcca gcaaggccac gccctctacc cccccaccct ctcccgggca ggatacattc 33660
ttggagctgg aagtttactg aatgggtccc ttgatcttga atgcactccc tgggccaagc 33720
atatgcaaag aaaaatgatg cttttatcac gtgtcctgct ctggctgcct ctgggttaag 33780
gataactttt gtacaggata caatcacaat gacatgcaca tcagggacat ttatggaaat 33840
attgtttttg ttctattcct ttccattttc aagaggatag tcttgtgaca tgttcaactg 33900
cctatgagag cctgtgatgg gcaggggtac tgtcctccac tgtgaggcac agggttagaa 33960
catggggcct gtgaggcggg cacatttgga gaatctactg gagaatgcca gagtgcattc 34020
aagatcaagg gacccattca gtaaacgtcc agctccaaga atgtattctg cccggtgtgt 34080
gtgtgtgtgt gtgtgtgtgt gtgtgtgaga gagagagaga gagagagaga gagagagaga 34140
gagagagaga gagagtgatg ggggacattc cagatcccct ttgagcctct tcactctctc 34200
tgtccagctc ctgtttggct gaagagcagc tgtggttcct cctgtggaag gaggaaagga 34260
gctagcatct ctggaaactg ctgcctactt tctagtctgc cagccccctg tgtattatta 34320
ctagctggtc tttacaaggg tccctggaag gtcaggaagt atgtgaaatc tggttaaaga 34380
agctctgcct cagaacagag ggtgaatctc agctattcca tgttaaatga caccttgtca 34440
ttagctgtcc tgtgcatgtt actggggagg aaagtgtttc tcagtcctag ccagctgtaa 34500
acctggcaag atatgtacac aggcgtaagg gagacatgaa tgttatgggg ccaaccaacc 34560
attttctaat ttttgtaaac aagtctttta atttgtttat ttagtttgta aatgtgtatg 34620
cacactcaac ttgcatttat gtttttgcac tgtgtgtatc cctggtgccc gtgaatgcca 34680
gacaaaaagt gtcaggtctc ttggaactgg agtacaggct gttgtgaacc accatgtgaa 34740
tttctcttca aggaacagca atgttctttt ttttttttt ttttaagatt tatttattac 34800
atgtaagtac actgtagctg tcttcagaca caccagaaga gggcgtcaga tctcgttaca 34860
gatggttgtg agccaccatg tggttgctgg gatttgaact ctggaccttc ggaagagcag 34920
tcgggtgctc ttacccactg agccatctca ccagcccaac agcaatgttc ttgactgctg 34980
aacctttttct ccagcctccc acaaaccact ttctggttgg acttgaagcc agctccacaa 35040
ggtagaaccc atgcctggta ccattaacgg ggctaaaaac tgtggctaac tacattatag 35100
gccatgggga gaacctagtc ttattatgtt aaatggacat agtaaaagac ttcccttcaa 35160
gatctcatct tcatactcag agatcagttc attcctcaac ttttatcaga gaaccgtctt 35220

tttgagtagg tggtgattga tacagcaact ggtcaccatg cagagactaa gactgtggac 35280
agctcagctc taaataggat gtctatgtca tgtgtcctcc ccacaaggct cagagaggga 35340
agagagggca gaaagtctgt aagagacaga gggagtgaac caatgcagtg agactgtgtt 35400
tgccagacac gataggacca ttgcatatgt gaactcacag gggctgggaa ggcatgtaca 35460
agacctgctt gcctaagatc aagccaacca caaccatagc aagataagtg agggcccaag 35520
aagtcccacc ccatctgagg cactactgac agctgagggc tacataatct cacccttctc 35580
cagggatgca ggctgtggtg gtttgactag gatcagctcc cacagactca tgtgtttgaa 35640
tgctttgctc ataaggagtg gcactattag gaggtgtggc cttgtcggag taggtgtggt 35700
tttgaggtct ttgcttaagc catacctagt gtggctcaca gtcacttcag ctgcctcttg 35760
atcaagatgg agaactctca gctccttctc cagaaccatg cctgcatgca tgctgccatg 35820
cttcccacca tgacaataat agactaaacc tctgaaatga caagtttggc tccaattaaa 35880
tgttttcctt ataagagttg ctgtggttat gatgtttcct taaagcaata gaaacccaca 35940
ttaagagacg gtccctgaga ggctacccat gttccagtaa acaggtccac actgaatgga 36000
tgaatggaca gcaatgaatg gactcagtgg gcatcaaatt gaaaagaaag ggggtggggt 36060
agaaaacatt tgtacatgaa gttgggaggg aaaaattgtg tggagctagg gaggatctgg 36120
aagggagagg atagggggta gattgaatcc aaacacacta tataatattt acgaatcata 36180
aaactaaaca acctcaagaa gaacctaggg gagcctgtat aatctgggag ggacagtttc 36240
cccaagcata gatccagaag ccgtaaacta aaagcaaggg ggccctgggg aagtggggaa 36300
gggaaaagac ccacaccccg ccagagttcc acctactctc tggtcagtca ggtgtgggag 36360
gggtgggcat tcctctatcc cactctttag ggagtggcca ggggcagccc tacctgggga 36420
ccctggagct actttgctaa agccaccagg gttataggag agagggatga gggaagagat 36480
tcccaacacc tgtgagagta catgcagcct tgatggagca gagactctct atggtttaag 36540
agctttatta tagaaaggca gggagagagg ggggggctag aaagagtaag agggagagag 36600
gagagaagtc aaagagagag gagagaggag aagacaaaga gagggtgaga gagagggtga 36660
gagtgagggg taagaagaag acaagtaaga ggagtaagag agcgaggtgg ggctgaacag 36720
ccctttttat ggtcttcact gttgctaggt aactggggag gagtttagtc tgaaggtcag 36780
aagcttgggc cattgcctaa gtgactactg accatgcttc tcttgttggg gctgtggggg 36840
acagtagctt aggcaggagc cagagttcca ggagcataag ggaacgccta ccgtgtcatg 36900
aaggtgaatt atgactttgg ggttcagaac tcagcttaac tggagaccag cctatctttg 36960
tatagcccaa tgccccacgc atattcaaat aggaaccctc tgtaatacaa accaaaataa 37020
acaaaccaaa atccaaaagt aagctggaaa atgcatagtg gtctgaaaaa acgtttgccc 37080
tacgttcaac atacgaggac taaaatcaag aatacacaaa gagacaaatt agtgcagtga 37140

gtatgtggat aactgtccca tgggaaacaa catctctggc taaaggaggc tggggaggtt 37200
gctcagttac taaagtgctt cctgttcaaa catgaggagc tgagttcaga tcctcagcat 37260
ccatggaaaa agcctgtgtg acagcatatg cttatgatcc cagtgctcca gaggcagagg 37320
aagaggatcc cagggcttgc tctcattcag tggagccaaa tccaagtgca gttgaaagac 37380
ctgcccccta ctccccaaca aaccaaagcc aaaacaaaat aaaaaccaag ctgggcagtg 37440
gtggcacatg cttttaatcc caacacttgg gaggctgaag caggcggatc tctgagtttg 37500
aggccatcct ggtctacaga gtgagttcca ggacagccag ggctacacag agaaacgctg 37560
tctccaaacc aaaccaacca actaaccaac caaccaacca accaaccaac caaccaacca 37620
accaaccagt ggagaataat tgagagagac acctgatgtg acttctggcc tccatctgta 37680
tgtgggaatg cacagtatca caaatgtatt tatttaacac acacacacat acacacacac 37740
acacagcagt taataaaaaa gataagctcc ccttttactg ttgcttgata gctcattcct 37800
tcttgttgct aaatagtatt tcctttcatg aaccattctc ttggcaaact cttggctgct 37860
tctaattttt gcagttatga ggaaggcaat gaaggtttct ttgtaggttt ttgtatgatc 37920
acagttttca aatgctgggc aaatatatgg tagcatgttt gctatgctgt aaagttacct 37980
ttagcttctc agtttttta ggtccttcct tctttctttt cctttctctt tccttccttc 38040
cttccttcct tccttccttc cttccttcct ttcttccttc ctggtggtaa gtgggactca 38100
ctttgtagct caggcttgtc ttgatactct tcctgtctca gccttccaag tgctggaact 38160
ttaaccataa gccaccccac cagactacta actattattt attggtttgt ttaattattg 38220
ctttttttct tttcttttag acaaataact cactgtattt agcctcagtg ggcctgccac 38280
ttgctgtata gacctggctg gccttgaact cacagaaatc agcctgtctc tgcctcctaa 38340
atactagagt taaacctgtg tgccaccatt cccagcttcc actaatttta tttacttttt 38400
ttttttttt ttttccgaga cagggtttct ctgtgtagcc ctggctgtcc tggaactcac 38460
tttgtagacc aggctggcct ctgcctccca agtgctggga ttaaaggctt gtgccacccc 38520
tgcccggcac tttatttact ttttgagtgt gtaattaatg caaatgcatt cagtagtacc 38580
ctttccttct attttgtacc ctatttctcc ttccttgatg tccctattcc agaggcagac 38640
tctgttcttt ctctcttttt tgtttttgtt ggtttgtttt ggagagaggg tgtcatgtga 38700
ttcagtctgg ccttaaagtc tctgtgtagc tactgttggc attcacgttc taatccttct 38760
gcctctgcat acaaatgcta gcatgccagg tgtgtaccac tatggctgat ttctgctctc 38820
ttccctgtga taccgtgtag atagtaaaga attattcaaa gtggctggga agattcctcg 38880
gtggttaaag cacttgccat gcaagtgtga ggactagaac ttggatcccc aagaaccaat 38940
caatgatcaa tgggcgtggt tgcctatact tccagcctca gcagagaaag ccggctggca 39000

tgaccagcta aagcagcgaa ctttgtattt gactgagaga cccttcctca atgaatggta 39060
gaagagtggg caaaggtgat tcctgacgtt agcaagtgct cttcaccaga aagccttttc 39120
tttagcattc acattatttc tttttaaaag tctgttgaca gcaagcagca ctgattcagt 39180
gaattacata aaaaaagtaa atgaggtcga aggggctcat gttgggggtt tggaatgtag 39240
gaattgtggt tcatatggtc aagatacatc gtatatatgc attaaattgt gaaaaaatat 39300
tcattttata tttttggttg tgagcctagc ctttaatggc tgagccatct ctccagccca 39360
aagatattct ttttttttt tgttttttt tgtttttttg agacagaaga tattcttttt 39420
taaaatatgt tgcctgttga ggcctgctcc ttttaatata gcagtagcca ttttgtattc 39480
tgtctccatt ttgctcctaa ggtgaaatga agttcaggtt ctcagactct gcttcccaga 39540
agtgagcatc cagagctgac actaagtatg ttactaataa gccaaaaagt tacggccgaa 39600
tcacttgtcc ctgttatctc aatgttctga aattccctgc tcagtacctg tccaccaccc 39660
ttcttacctc agtcaggacc actcagctta caggttggct aataatactt tatctagtta 39720
gacaaaactg ctgtaccact tcactgcttg cctttgaacc tttttttaag atatatttat 39780
tatttatatg taagtacact gtagctgtct tcagacgcac cagaaaaggg tgtcaaatct 39840
tattacggat ggttgtaagc caccatgtgg ttgctgggat ttgaactcag gaccttccga 39900
agagcagtca gtgctcttaa ccgctgagcc atctctccag ccccatcttt gaacttttga 39960
acctggtttt tcctataaaa agcctgccct gaggaccggc tggtgccaca gttaggtttt 40020
tccttcttgt ggacctagat gtccagtatt atgctgtgtg ttcaataaac tattcctgtt 40080
taactgaaat tggtgtacgt atggtttgtg gcaagtctca gaccccgaca ctgacatgtg 40140
atgtgtatgg ttatttgcaa attaataaat ttaagcatta actttcagta tagtaaataa 40200
tgatgaataa aacataaaaa ctgtttggat tgtcaacaaa attttctcat cgtctgtgta 40260
tgggtgtttt gcctctaggc atgtatgtac ttcatatgtg catgcagtgt cctctaacgt 40320
cagaagaggg tagcagattc cctgggttta tagatgattg tgagccacca tgtgggtgct 40380
gggaatcaaa tctgggtcct ctggaatatc agctagtgtt ctttgttttt gtttttctga 40440
gacaggattt ctccatgtag ccctggctgt cctgaactct cttggtagac caggctggcc 40500
ttgaattcat agagatctgc ctacctcact ggcattaaag gtgtgtgcca ccaccaactg 40560
tctgagccat cgcggttgct ccatgggttc ttgaaacaaa atttaagatc ataatttttt 40620
gtttggttgg ttttttttga cacagggttt ctctgtgtag ctctggccgt cctggaactt 40680
actctgtaga ttaggctggc ctcgaactca gaaatccgcc tgcctctgcc ttccaagagc 40740
acgatcataa attctaagtt gaaaaaattt acatcaattt atctgtatgg ctttacttaa 40800
aatttgctaa ggcccaacac tattaagtta tttgttaagc ctggaatgtc tcttactgga 40860
aaagcatttt cctaacatgt tcaagaccct gagtttgtct cctagaacta caagaaaaca 40920

aaagtaaaat cagtattttg ctctgtgtgg tggcatatat ctttacttct gcacttggca 40980
ggcagaaaca ggcatatttc tgtgaatttg aggacagttt ggtctataca gcgagttcca 41040
gggcagccaa ggctgcagag taagacaatg tcttttaaaa aagttaattt gtagatgtta 41100
gttagtctag tagagatgaa attcattaaa atctttttttg ttgttgttgt tgttttttt 41160
ccggtcagga tctccttgca gagcccaagt tggtcttgaa ctggctatgt ggatgaataa 41220
acatttctgg ttctgtcgcc accttcccag tgctaggttt acaggtatgg gttactacac 41280
agtttataca acattcagga cacattagtc acacatgtgg gttactacac agtttataca 41340
acactcagga cacattagtc acacacttta ccaatttagc tacattgaat gaaaaaaaaa 41400
caaaaaggag gacatgatgg cttctaggta tggtggagga ggtttattgt agacaggagg 41460
gagcagacag ccagaagcag aggcatctgg gagaattcag ggtggaagtg gctgtagaat 41520
gagctgggcc atgtgagaag ggttagggga gagggtagaa gagacctgga gtcaagaggc 41580
caggagacca agaggccaaa gggtaaaaag gacctcataa ccaaaatggc tgggttacat 41640
aggaatcaga gaagcttggg gaggaaaagg ccagctcaga ctctggactg gagaagttta 41700
gggtagaggt caggattagt atgccagcca gaaggatcct gtaccagaag gtactgaggg 41760
agactggtgg ccagagtctg ctttgatatt ttattaggca tctcagccat ttgtcttcgg 41820
tttgtgacct agcatatgtt cctaatctgt ctgtctttcg tctcccccca accctcttct 41880
gagactgggt ttgactctat agcccaggtt ggccatgatt tggatgcggg gattacaggt 41940
ataaaccaca gattgattgt gtgtcaacct tacataaacg ttttcttaaa atgtctatat 42000
gcacgtatta gtaacggcac gtgtatacac acatgctata catacggatg cacacacaca 42060
tatttacagt atcatacttc ttttttcct tcactgagct tttcacactt ttcttgtcct 42120
ttcaagtcac cttgaaatct ggtgcaacct tctgagtttc actcatagct ctgttaatag 42180
gaattccata caattctaca atattccttt tccgttccct ccgctaacaa acaaaatgtg 42240
gtattataag aaggcgcacc agacacctac gcaggaattc aatccagaaa gaagaaaggc 42300
tcccagacca cgtgacacct ccagggacta cgtcaagtgg ccgtcaccac aatgcttccg 42360
ccctcttcaa acatggttgg caagcgctct ccgcatcgtg accatggtta ttcttgcatg 42420
taggaaccgt actgagcgca taccaatctc ctttaggcaa gtgtcgcggc ggaggagatc 42480
cagcagagcc gcaagaacga cgatcggtta ccgccggtag taacaagcgc ggaccggaag 42540
ttccgcgtct tcgctgtccg gggggagccg ttaggcgcgc acgccggaag tggccaatca 42600
gccggtgtga ggcggtgccc actgtgttcg cgtccctcgg gcagcagagc catggagccc 42660
ggggctgctg agctttatga ccaggccctg ttgggcatcc tgcagcacgt gggcaatgtc 42720
caggactttc tgcgcgtgct cttcggcttt ctctaccgca agaccgactt ctaccgcctg 42780

ctgcgccacc cttcggaccg catgggcttc ccgcccgggg ccgcacaggc cctggtgctg 42840
caggtgaggt ggagagaggc ggcgggccgt tggggtccag caggtcctta ccccagttcc 42900
acctcccagc gccagaggtg ccccggctcg cgtcctacgt ctgggaactg cgccactccg 42960
tagcccaccc ttcagggttt gtgattctct ctggggtgcc accaggtgat ttgagtaaat 43020
ggccaggcgt tatctgaccc acggatccgt tggcaggaat gcgcttcttt gggtacaggc 43080
tgggtttgtg cgggagatgc ttaggtgttg gacctgtcag cctgggtttg agggcctccc 43140
agcgccctgt gggcatcccc agaacggtgc aaggggcttg ttaggcaggg ttcaaaccgt 43200
gcaggctgta ggaggaggac tttcacagcg ggagaaacta gtagatttca gaattcccgg 43260
gctcggaggt ggcaggaatg actggtatgt attgattggt gggtgtggcc tcttgccagt 43320
gacttactaa gttggtggac aattgtaatg tttacgaaat ttacatttgg attaaaattt 43380
atatgtccta ggttatattt attgtttttg aagcgttact aatgactgac tcaagttgtt 43440
tggacacctt tttaaaaact gttttctttc gagacaagat cttgctgtgt agccctggtt 43500
gacttggaac ttgccaggta gatcaggatg gcttcgaact tatggtcctc cttcccgacg 43560
ctcccgcgct gggattgttg acattgttga gttagtaata agcaaacatt taatgaatat 43620
atttaacact ctcagctgcg gtatagactc tgatggtatg gatatttaaa taatggtctg 43680
tctacttttg acactgtaca gagagtaaag cacagacact aatagagagc tctgctagtt 43740
ctctgtgtgg tagagctctt tggaaggaag tgatagaagt ggtagttagg aaattgagac 43800
agttttgtag gagacgttga agtaggtttt gaggggcgtg tagagtctga ttcacaaacc 43860
tataaagtgc ttcatgtcac ctttgctgtt tgttgacccc tgcatcagcc ccagacaggg 43920
ctttcctgtg tagctctgaa gtcctggaac tcactctggc atcaacctca gatatctgcc 43980
taggattaaa gactcgaatt gccaccacct ggcttacctt ccatcttaat ctgttctgtg 44040
ggcttgtcag gggttacttt cttcagttcc tctcagtgga agaagccaaa gcttagaatg 44100
ttctagtaac tcttaacata atgaccatga ggtgatcagt gggaaggttt gctactgagt 44160
ttcagtttca gaatggaaac cgatgagtcc ctggggtagc tcatccttat cagtggccca 44220
agtgctgctg gttgagtaat tggagagggt gagaggtggc accttgtgtt tctttataat 44280
gaagtcttgc cacctgatcg ccatgctctc tgaatctgac gtagtagttc aacaaggtta 44340
gatgacaaca agaacttcac tctgtttcct gcaactcagg acttcagttt tctaattcca 44400
aatctttttg ccctttttgt catttagcca ctccatagta tggtgagaac ttttgttttg 44460
gttatgaaag gaggaaagag tatcccagtg gctggctggc atttgaattt cttctgtata 44520
acatcttatt ttagctttaa cacaatttga gaggttggtt cttgtttgtc tgatgaactg 44580
ataaaggcaa gatagatgca cttatcgtac aactttataa aacagctacc tctgaaaggt 44640
taagatagct ctataggtta cagtgtggtg cccatgctat gggaggtagg gagccgagga 44700

agggacaggg tcaaccaaat atttcaaggt tccaaatctg ggtaactgga agagtgaggt 44760
cattaactgt attgtcactt ggatttgggg ggtgggaaag ttttagataa gttgaatttg 44820
agatcgcaga caggttctgg cttcagagct gttggaatgg taggactgga tcttgcagga 44880
gatggaagga ctgtgaagtc tgaagaagag cccagtgtag aggacagctg gggaacggtc 44940
accgtaaggg gcattgggtt ctaccgcagc caaggttaca gtcccaagta caggtttaag 45000
agttgtttgt tcttaccctt tctccaaccc caggtggttc tcaaccatca gccttgaccg 45060
agagcccttt tatcttctca gtttctttct tttgtggggt gggaggttga agacaaggtt 45120
tctctgtgta gctttggcta tcctgaaact tgctttgtag tccaagctgg ccttgaactc 45180
acagagatcc gcttgcctct gccttctgag cgccggcatt aaataaaggc atgtgccacc 45240
actgcctggc tcagtttcta ttttcaaaca agtatttatt gagtctccac tataatatac 45300
actgatttgg gccatgagaa acagaggctt ataagttgtt gtggggtttt ggatttttt 45360
tttattctga gaaaaggtct caccatgtag ccttgactgg cttggaacct actatgtaga 45420
tcaggctggt ctcgtattca aagacatctg cctgcctctg cttcttgagt gctgggacta 45480
aaggcgtgcg ccaccacatc cagccaacca agagactttt tactatgtat acacttgaat 45540
actattttca tcaggtagat catagtaagt cccccgagga tctgcttttt ttttttttt 45600
tttaatttac tgaaagcctt tgcaagaggc ttgtaagcta catttagtat tggtaagggt 45660
tttggtatct tttctactca tttactgctt tttctgttct tgattcagtg atcctaagac 45720
tccttcctgt gacttcctct ctcaaagtgt ggttgctgcc cagggtctgt gaacaggccg 45780
gagtggtttg ggtggtttct gtggcgcttg ggagctccct accgatgtct tggggagtag 45840
ctgctgctgt gtgtgcttct ttatgtcaac tgaatccaaa ccgagaccaa ggcgtctaga 45900
aagacataat ctggggctgg tgagatggct cagtgggtaa gagcacccga ctgctcttcc 45960
gaaggtccga agttcaaatc cctgcaacca catggtggct cacaaccatc cacaacaaga 46020
tctgactccc tcttctggag ggtctgaaga cagctacagt gtacttacat gtaataaata 46080
aataaataaa cctttaaaaa aaaaaaagac ataatctcag cccaggactt gccttcatca 46140
gatgattgat gtgggagggc ccactgtggg cagtgccatc cctgggcagg tggtatgggg 46200
tgtataagaa agcaggctgc gtgagttagt aggtgtcatt cctccgtagt ctctgcttta 46260
gttcttgctt caagttcctg ccttggcttc acctgatgat ggactggatc ctttaagcca 46320
tgtaaaccct ttcctcccca ggttgtttgt gatcatggtg tgttttacac agccacagaa 46380
agcacacaag tgcagtgctg ctgacagagg ctggctgtct gtttcccttc tcagggttta 46440
gagcatgagt gaagtgagta aagattttgt tctgttcatt agtattagaa aatgtttttg 46500
ttttgttttg ttttttttg ttttttgttt ttcgagacag ggtttctctg tacagccctg 46560

gctgtcctgg aactcactct gtagaccagg ctggcctcga actcagaaat ccacctgccc 46620
ctgcccctgc ccccccccc ccccccccagt gctgggatta aaggcgtgca ctgccacgcc 46680
cggctagtat tagaaaatgt taagaacaaa agtagtctag tcagtcaatg actaaacaga 46740
caatgactcc tggagtcagt gtgaaaagca tggctaagtg gaagtcttcc ttataggagt 46800
caggagccac ctgtttctca cttctcttgt aacttattgg aaaatcttag aggagtcagt 46860
tcccctcttg tcttctgaga tggcatattg aaggcaatgg acttctattc caacaaggac 46920
actgcctcta gggtgagtct gtgaacatag gcagctggag ctctggagtg ttgtgagaac 46980
tgggtagtgt agatgggcag gtggaagggg agtcgcagga aggctgagtg aggactgtca 47040
ggccttggtt tggagcaccc tgaggttaca ggagagtctt tactgcctga cttcttgtag 47100
ttaagttgaa gatttaaggt gttagagata gagtgctttt gagtacaaag gactcagaca 47160
ctggtagcat ggatgctaga gagagtcaaa ccttcaagct accagcagat tagaagtggt 47220
ttttgactgt gttttgtttt gttgttgttg tcgtcgtcct tgtcgtcgtc atcatcttct 47280
tcttctaaag atttatttat ttatttttatg tgtatgagta cactgtagct gtacagatgg 47340
ttgtgagcca tcatgtggtt gctgaaaatt gaactcagga cctctgcttg ctccagcccc 47400
acttgatcct gccccgcttg ctcctgtcct aagatttatt atatgtaagc acactgtagc 47460
tgtcttcagc cacaccagaa gagggtgtca gatctcatta cggatggttg tgaaccacca 47520
tgtggttgct gggatttgaa ctctgctctt ccgaaggtct tgagtgctct gagccatctc 47580
tccagccctg tctttttttt ttttttttga aacagggtct gactgtacac cctggctggg 47640
ctggaactca ctatatatca atcacaattt atatcaggct agcctctgcc tcctgggtgc 47700
tgaattaaag atgtgtgcta ccatacctgg cctttgtctc cagttttaca tcttttagga 47760
ttctgtctgt ctgtctgtcc atttatttttg gctagattga cttttattgt ttgttgtgtc 47820
ttgtgtaact tctctgaatc tgcattttct tcatttgagg atgttggtgg tgctcttcac 47880
agggtttttg tgagttttga aactgtagaa gcacacagta tagctagcac tgtgttttgt 47940
cttccgagtt gtgctcagac atgttagtga gtactcagtg ccccagccat gtcccagctt 48000
acttcctcaa gccttattac ctctgcctag tgcaggggtc ttgctctttc ttggtgatac 48060
tgcttcttgg agctctgctg ggcacactcc ttcattaagg accacttgag aacccgatcc 48120
ttcttcctct gtagatttct ttctctagaa atgttctcac cacagtctgc cagggcatca 48180
gggttccatg atgccccacc tgtatagact tctatcagag tgagtctgta cttgtgcagt 48240
gttgaagata aaacccactc tcctgtcatg accagccaag ctgacctaca ccccaagccc 48300
tatagctgta tggaccatcc atgatagtct cagcaccttt ctggatgtta agggtgtgtc 48360
ttcataccaa aggaggactg tcaggcccca gtcttggaaa agcctgtgtt agaagtccca 48420
ggtaatagga gtgagtttgt attcttttgc tttttagagt ttttattcct tgcacactgt 48480

aggcccaggg tgggtgttta tggagtcaag tactcactcc tacatcgtat cctcagctgt 48540
cagttgggga agtggtggca agaatctgat aagcctgagt gcatcttgta gattttcatc 48600
tttcactttt taaacctgaa ttgctggcac cttccggaat ccacagcctg agtgtgttct 48660
tcacattgcc agcaaggttg gcaaaagtaa tgacaactct ggtgtcggcc tttaatctca 48720
gcactgggaa gacagaggca ggcagatcgc tgaggccagc ctggtctaca gagcaaattc 48780
caggacatcc agggctacac agagaaactc tgtcttgcta caacctcccc cttctcctac 48840
cttggtccca aaaagtaatg acaactaaag ctgtctagtt ttgcatcctc taggtttgta 48900
catacagtta gatttgactt attttgagtt tattcatttt ggaaacttct tgaggaagag 48960
caattcctac cagcttttgg tgaagttgta gagtgtttcc attttgcttt ggtttactgt 49020
tatttaattt tatacttaga agattttgtt atatttctgt ggtggtgtgg tctgataggg 49080
tgcagatgaa tttatttatt tatttattta tttatttatt tatttattta tttatttttg 49140
agacagggtt tctctgtgta gccctggctg tcctggaact cactctgtag accaggctgg 49200
cctcgaactc agaaatccat ctgcctctgc ctcctgagtg ctgggattaa aggtgtgagc 49260
caacactgcc cagctgcaga tgttgtattg atgtttgttt catttttagg tctttaaaac 49320
atttgatcac atggcccgcc aggatgatga gaaaaggaag aaagaactag aagagaaaat 49380
aagaaaaaag gaggaagagg ccaaggcctt gccagctgct gaaactgaga aggtagcggt 49440
gccggtccca gtgcaggagg tagagatcga tgctgctgca gacttgagtg ggcctcagga 49500
agtagagaag gaggagcccc caggctccca ggaccccgag cacacagtga cccatggcct 49560
ggagaaggcg gaagctccag gaacagttag cagtgctgct gaaggcccta aggaccctcc 49620
tgtgctcccc aggtaggagc atctcctgca gtgtcgtcct ctctgctgtg cttaagtttg 49680
cctatgagtg gtttttgttt tgtgtggttt gtaaaaaaat atcagctctg ttttggtggg 49740
cagtggttaa ctatagaaat tcatttctta atagttctgg aggctggaaa ccccagatta 49800
aagtatgatc tgggttgttt gtttgaaggc ttctatcagt ggtttgcaga cagccatctt 49860
cctgtgtctc gtcacattcc tttgttcttg cctgtgtctt attctcctat tctcaagcag 49920
cactcaaaca ccttagtgag ccagattgcc ttccatcctg gtgcttcagg gacctccttt 49980
cggaactcag tgttgaattc 50000

<210> 2
<211> 4002
<212> DNA
<213> Mus musculus

<400> 2
atggcagctg cctggcaggg atggctgctc tgggccctgc tcctgaattc ggcccagggt 60
gagctctaca cacccactca caaagctggc ttctgcacct tttatgaaga gtgtgggaag 120

aacccagagc tttctggagg cctcacatca ctatccaata tctcctgctt gtctaatacc 180
ccagcccgcc atgtcacagg tgaccacctg gctcttctcc agcgcgtctg tccccgccta 240
tacaatggcc ccaatgacac ctatgcctgt tgctctacca agcagctggt gtcattagac 300
agtagcctgt ctatcaccaa ggccctcctt acacgctgcc cggcatgctc tgaaaatttt 360
gtgagcatac actgtcataa tacctgcagc cctgaccaga gcctcttcat caatgttact 420
cgcgtggttc agcgggaccc tggacagctt cctgctgtgg tggcctatga ggccttttat 480
caacgcagtt ttgcagagaa ggcctatgag tcctgtagcc gggtgcgcat ccctgcagct 540
gcctcgctgg ctgtgggcag catgtgtgga gtgtatggct ctgccctctg caatgctcag 600
cgctggctca acttccaagg agacacaggg aatggcctgg ctccgctgga catcaccttc 660
cacctcttgg agcctggcca ggccctggca gatgggatga agccactgga tgggaagatc 720
acaccctgca atgagtccca gggtgaagac tcggcagcct gttcctgcca ggactgtgca 780
gcatcctgcc ctgtcatccc tccgcccccg gccctgcgcc cttctttcta catgggtcga 840
atgccaggct ggctggctct catcatcatc ttcactgctg tctttgtatt gctctctgtt 900
gtccttgtgt atctccgagt ggcttccaac aggaacaaga acaagacagc aggctcccag 960
gaagccccca acctccctcg taagcgcaga ttctcacctc acactgtcct tggccggttc 1020
ttcgagagct ggggaacaag ggtggcctca tggccactca ctgtcttggc actgtccttc 1080
atagttgtga tagccttgtc agtaggcctg acctttatag aactcaccac agaccctgtg 1140
gaactgtggt cggcccctaa aagccaagcc cggaaagaaa aggctttcca tgacgagcat 1200
tttggcccct tcttccgaac caaccagatt tttgtgacag ctaagaacag gtccagctac 1260
aagtacgact ccctgctgct agggcccaag aacttcagtg ggatcctatc cctggacttg 1320
ctgcaggagc tgttggagct acaggagaga cttcgacacc tgcaagtgtg gtcccatgag 1380
gcacagcgca acatctccct ccaggacatc tgctatgctc ccctcaaccc gcataacacc 1440
agcctcactg actgctgtgt caacagcctc cttcaatact tccagaacaa ccacacactc 1500
ctgctgctca cagccaatca gactctgaat ggccagacct ccctggtgga ctggaaggac 1560
catttcctct actgtgccaa tgcccctctc acgtacaaag atggcacagc cctggccctg 1620
agctgcatag ctgactacgg ggcacctgtc ttccccttcc ttgctgttgg gggctaccaa 1680
gggacggact actcggaggc agaagccctg atcataacct tctctatcaa taactacccc 1740
gctgatgatc cccgcatggc ccacgccaag ctctgggagg aggctttctt gaaggaaatg 1800
caatccttcc agagaagcac agctgacaag ttccagattg cgttctcagc tgagcgttct 1860
ctggaggacg agatcaatcg cactaccatc caggacctgc ctgtctttgc catcagctac 1920
cttatcgtct tcctgtacat ctccctggcc ctgggcagct actccagatg gagccgagtt 1980

gcggtggatt ccaaggctac tctgggccta ggtggggtgg ctgttgtgct gggagcagtc 2040
gtcgctgcca tgggcttcta ctcctacctg ggtgtcccct cctctctggt catcattcaa 2100
gtggtacctt tcctggtgct ggctgtggga gctgacaaca tcttcatctt tgttcttgag 2160
taccagaggc tgcctaggat gcccggggag cagcgagagg ctcacattgg ccgcaccctg 2220
ggtagtgtgg cccccagcat gctgctgtgc agcctctctg aggccatctg cttctttcta 2280
ggggccctga cctccatgcc agctgtgagg acctttgcct tgacctctgg cttagcaatc 2340
atctttgact tcctgctcca gatgacagcc tttgtggccc tgctctccct ggatagcaag 2400
aggcaggagg cctctcgccc cgacgtcgtg tgctgctttt caagccgaaa tctgccccca 2460
ccgaaacaaa aagaaggcct cttactttgc ttcttccgca agatatacac tcccttcctg 2520
ctgcacagat tcatccgccc tgttgtgctg ctgctctttc tggtcctgtt tggagcaaac 2580
ctctacttaa tgtgcaacat cagcgtgggg ctggaccagg atctggctct gcccaaggat 2640
tcctacctga tagactactt cctctttctg aaccggtact tggaagtggg gcctccagtg 2700
tactttgaca ccacctcagg ctacaacttt tccaccgagg caggcatgaa cgccatttgc 2760
tctagtgcag gctgtgagag cttctcccta acccagaaaa tccagtatgc cagtgaattc 2820
cctaatcagt cttatgtggc tattgctgca tcctcctggg tagatgactt catcgactgg 2880
ctgaccccat cctcctcctg ctgccgcatt tatacccgtg gcccccataa agatgagttc 2940
tgtccctcaa cggatacttc cttcaactgt ctcaaaaact gcatgaaccg cactctgggt 3000
cccgtgagac ccacaacaga acagtttcat aagtacctgc cctggttcct gaatgatacg 3060
cccaacatca gatgtcctaa agggggccta gcagcgtata gaacctctgt gaatttgagc 3120
tcagatggcc agattatagc ctcccagttc atggcctacc acaagccctt acggaactca 3180
caggacttta cagaagctct ccgggcatcc cggttgctag cagccaacat cacagctgaa 3240
ctacggaagg tgcctgggac agatcccaac tttgaggtct tcccttacac gatctccaat 3300
gtgttctacc agcaatacct gacggttctc cctgagggaa tcttcactct tgctctctgc 3360
ttcgtgccca cctttgtggt ctgctacctc ctactgggcc tggacatacg ctcaggcatc 3420
ctcaacctgc tctccatcat tatgatcctc gtggacacca tcggcctcat ggctgtgtgg 3480
ggtatcagct acaatgctgt gtccctcatc aaccttgtca cggcagtggg catgtctgtg 3540
gagttcgtgt cccacattac ccggtccttt gctgtaagca ccaagcctac ccggctggag 3600
agagccaaag atgctactat cttcatgggc agtgcggtgt ttgctggagt ggccatgacc 3660
aacttcccgg gcatcctcat cctgggcttt gctcaggccc agcttatcca gattttcttc 3720
ttccgcctca acctcctgat caccttgctg ggtctgctac acggcctggt cttcctgccc 3780
gttgtcctca gctatctggg gccagatgtt aaccaagctc tggtactgga ggagaaacta 3840
gccactgagg cagccatggt ctcagagcct tcttgcccac agtacccctt cccggctgat 3900

gcaaacacca gtgactatgt taactacggc tttaatccag aatttatccc tgaaattaat 3960

gctgctagca gctctctgcc caaaagtgac caaaagttct aa 4002

<210> 3
<211> 1333
<212> PRT
<213> Mus musculus

<400> 3

Met Ala Ala Ala Trp Gln Gly Trp Leu Leu Trp Ala Leu Leu Leu Asn
1 5 10 15

Ser Ala Gln Gly Glu Leu Tyr Thr Pro Thr His Lys Ala Gly Phe Cys
20 25 30

Thr Phe Tyr Glu Glu Cys Gly Lys Asn Pro Glu Leu Ser Gly Gly Leu
35 40 45

Thr Ser Leu Ser Asn Ile Ser Cys Leu Ser Asn Thr Pro Ala Arg His
50 55 60

Val Thr Gly Asp His Leu Ala Leu Leu Gln Arg Val Cys Pro Arg Leu
65 70 75 80

Tyr Asn Gly Pro Asn Asp Thr Tyr Ala Cys Cys Ser Thr Lys Gln Leu
85 90 95

Val Ser Leu Asp Ser Ser Leu Ser Ile Thr Lys Ala Leu Leu Thr Arg
100 105 110

Cys Pro Ala Cys Ser Glu Asn Phe Val Ser Ile His Cys His Asn Thr
115 120 125

Cys Ser Pro Asp Gln Ser Leu Phe Ile Asn Val Thr Arg Val Val Gln
130 135 140

Arg Asp Pro Gly Gln Leu Pro Ala Val Val Ala Tyr Glu Ala Phe Tyr
145 150 155 160

Gln Arg Ser Phe Ala Glu Lys Ala Tyr Glu Ser Cys Ser Arg Val Arg
165 170 175

Ile Pro Ala Ala Ala Ser Leu Ala Val Gly Ser Met Cys Gly Val Tyr
180 185 190

Gly Ser Ala Leu Cys Asn Ala Gln Arg Trp Leu Asn Phe Gln Gly Asp
195 200 205

Thr Gly Asn Gly Leu Ala Pro Leu Asp Ile Thr Phe His Leu Leu Glu
210 215 220

Pro Gly Gln Ala Leu Ala Asp Gly Met Lys Pro Leu Asp Gly Lys Ile
225 230 235 240

Thr Pro Cys Asn Glu Ser Gln Gly Glu Asp Ser Ala Ala Cys Ser Cys
245 250 255

Gln Asp Cys Ala Ala Ser Cys Pro Val Ile Pro Pro Pro Pro Ala Leu
260 265 270

Arg Pro Ser Phe Tyr Met Gly Arg Met Pro Gly Trp Leu Ala Leu Ile
275 280 285

Ile Ile Phe Thr Ala Val Phe Val Leu Leu Ser Val Val Leu Val Tyr
290 295 300

Leu Arg Val Ala Ser Asn Arg Asn Lys Asn Lys Thr Ala Gly Ser Gln
305 310 315 320

Glu Ala Pro Asn Leu Pro Arg Lys Arg Arg Phe Ser Pro His Thr Val
325 330 335

Leu Gly Arg Phe Phe Glu Ser Trp Gly Thr Arg Val Ala Ser Trp Pro
340 345 350

Leu Thr Val Leu Ala Leu Ser Phe Ile Val Val Ile Ala Leu Ser Val
355 360 365

Gly Leu Thr Phe Ile Glu Leu Thr Thr Asp Pro Val Glu Leu Trp Ser
370 375 380

Ala Pro Lys Ser Gln Ala Arg Lys Glu Lys Ala Phe His Asp Glu His
385 390 395 400

Phe Gly Pro Phe Phe Arg Thr Asn Gln Ile Phe Val Thr Ala Lys Asn
405 410 415

Arg Ser Ser Tyr Lys Tyr Asp Ser Leu Leu Leu Gly Pro Lys Asn Phe
420 425 430

Ser Gly Ile Leu Ser Leu Asp Leu Leu Gln Glu Leu Leu Glu Leu Gln
435 440 445

Glu Arg Leu Arg His Leu Gln Val Trp Ser His Glu Ala Gln Arg Asn

450 455 460

Ile Ser Leu Gln Asp Ile Cys Tyr Ala Pro Leu Asn Pro His Asn Thr
465 470 475 480

Ser Leu Thr Asp Cys Cys Val Asn Ser Leu Leu Gln Tyr Phe Gln Asn
485 490 495

Asn His Thr Leu Leu Leu Leu Thr Ala Asn Gln Thr Leu Asn Gly Gln
500 505 510

Thr Ser Leu Val Asp Trp Lys Asp His Phe Leu Tyr Cys Ala Asn Ala
515 520 525

Pro Leu Thr Tyr Lys Asp Gly Thr Ala Leu Ala Leu Ser Cys Ile Ala
530 535 540

Asp Tyr Gly Ala Pro Val Phe Pro Phe Leu Ala Val Gly Gly Tyr Gln
545 550 555 560

Gly Thr Asp Tyr Ser Glu Ala Glu Ala Leu Ile Ile Thr Phe Ser Ile
565 570 575

Asn Asn Tyr Pro Ala Asp Asp Pro Arg Met Ala His Ala Lys Leu Trp
580 585 590

Glu Glu Ala Phe Leu Lys Glu Met Gln Ser Phe Gln Arg Ser Thr Ala
595 600 605

Asp Lys Phe Gln Ile Ala Phe Ser Ala Glu Arg Ser Leu Glu Asp Glu
610 615 620

Ile Asn Arg Thr Thr Ile Gln Asp Leu Pro Val Phe Ala Ile Ser Tyr
625 630 635 640

Leu Ile Val Phe Leu Tyr Ile Ser Leu Ala Leu Gly Ser Tyr Ser Arg
645 650 655

Trp Ser Arg Val Ala Val Asp Ser Lys Ala Thr Leu Gly Leu Gly Gly
660 665 670

Val Ala Val Val Leu Gly Ala Val Val Ala Ala Met Gly Phe Tyr Ser
675 680 685

Tyr Leu Gly Val Pro Ser Ser Leu Val Ile Ile Gln Val Val Pro Phe
690 695 700

Leu Val Leu Ala Val Gly Ala Asp Asn Ile Phe Ile Phe Val Leu Glu
705 710 715 720

Tyr Gln Arg Leu Pro Arg Met Pro Gly Glu Gln Arg Glu Ala His Ile
725 730 735

Gly Arg Thr Leu Gly Ser Val Ala Pro Ser Met Leu Leu Cys Ser Leu
740 745 750

Ser Glu Ala Ile Cys Phe Phe Leu Gly Ala Leu Thr Ser Met Pro Ala
755 760 765

Val Arg Thr Phe Ala Leu Thr Ser Gly Leu Ala Ile Ile Phe Asp Phe
770 775 780

Leu Leu Gln Met Thr Ala Phe Val Ala Leu Leu Ser Leu Asp Ser Lys
785 790 795 800

Arg Gln Glu Ala Ser Arg Pro Asp Val Val Cys Cys Phe Ser Ser Arg
805 810 815

Asn Leu Pro Pro Pro Lys Gln Lys Glu Gly Leu Leu Leu Cys Phe Phe
820 825 830

Arg Lys Ile Tyr Thr Pro Phe Leu Leu His Arg Phe Ile Arg Pro Val
835 840 845

Val Leu Leu Leu Phe Leu Val Leu Phe Gly Ala Asn Leu Tyr Leu Met
850 855 860

Cys Asn Ile Ser Val Gly Leu Asp Gln Asp Leu Ala Leu Pro Lys Asp
865 870 875 880

Ser Tyr Leu Ile Asp Tyr Phe Leu Phe Leu Asn Arg Tyr Leu Glu Val
885 890 895

Gly Pro Pro Val Tyr Phe Asp Thr Thr Ser Gly Tyr Asn Phe Ser Thr
900 905 910

Glu Ala Gly Met Asn Ala Ile Cys Ser Ser Ala Gly Cys Glu Ser Phe
915 920 925

Ser Leu Thr Gln Lys Ile Gln Tyr Ala Ser Glu Phe Pro Asn Gln Ser
930 935 940

Tyr Val Ala Ile Ala Ala Ser Ser Trp Val Asp Asp Phe Ile Asp Trp
945 950 955 960

Leu Thr Pro Ser Ser Ser Cys Cys Arg Ile Tyr Thr Arg Gly Pro His
965 970 975

Lys Asp Glu Phe Cys Pro Ser Thr Asp Thr Ser Phe Asn Cys Leu Lys
980 985 990

Asn Cys Met Asn Arg Thr Leu Gly Pro Val Arg Pro Thr Thr Glu Gln
995 1000 1005

Phe His Lys Tyr Leu Pro Trp Phe Leu Asn Asp Thr Pro Asn Ile
1010 1015 1020

Arg Cys Pro Lys Gly Gly Leu Ala Ala Tyr Arg Thr Ser Val Asn
1025 1030 1035

Leu Ser Ser Asp Gly Gln Ile Ile Ala Ser Gln Phe Met Ala Tyr
1040 1045 1050

His Lys Pro Leu Arg Asn Ser Gln Asp Phe Thr Glu Ala Leu Arg
1055 1060 1065

Ala Ser Arg Leu Leu Ala Ala Asn Ile Thr Ala Glu Leu Arg Lys
1070 1075 1080

Val Pro Gly Thr Asp Pro Asn Phe Glu Val Phe Pro Tyr Thr Ile
1085 1090 1095

Ser Asn Val Phe Tyr Gln Gln Tyr Leu Thr Val Leu Pro Glu Gly
1100 1105 1110

Ile Phe Thr Leu Ala Leu Cys Phe Val Pro Thr Phe Val Val Cys
1115 1120 1125

Tyr Leu Leu Leu Gly Leu Asp Ile Arg Ser Gly Ile Leu Asn Leu
1130 1135 1140

Leu Ser Ile Ile Met Ile Leu Val Asp Thr Ile Gly Leu Met Ala
1145 1150 1155

Val Trp Gly Ile Ser Tyr Asn Ala Val Ser Leu Ile Asn Leu Val
1160 1165 1170

Thr Ala Val Gly Met Ser Val Glu Phe Val Ser His Ile Thr Arg
1175 1180 1185

Ser Phe Ala Val Ser Thr Lys Pro Thr Arg Leu Glu Arg Ala Lys
1190 1195 1200

Asp Ala Thr Ile Phe Met Gly Ser Ala Val Phe Ala Gly Val Ala
1205 1210 1215

Met Thr Asn Phe Pro Gly Ile Leu Ile Leu Gly Phe Ala Gln Ala
1220 1225 1230

Gln Leu Ile Gln Ile Phe Phe Phe Arg Leu Asn Leu Leu Ile Thr
1235 1240 1245

Leu Leu Gly Leu Leu His Gly Leu Val Phe Leu Pro Val Val Leu
1250 1255 1260

Ser Tyr Leu Gly Pro Asp Val Asn Gln Ala Leu Val Leu Glu Glu
1265 1270 1275

Lys Leu Ala Thr Glu Ala Ala Met Val Ser Glu Pro Ser Cys Pro
1280 1285 1290

Gln Tyr Pro Phe Pro Ala Asp Ala Asn Thr Ser Asp Tyr Val Asn
1295 1300 1305

Tyr Gly Phe Asn Pro Glu Phe Ile Pro Glu Ile Asn Ala Ala Ser
1310 1315 1320

Ser Ser Leu Pro Lys Ser Asp Gln Lys Phe
1325 1330

<210> 4
<211> 30
<212> DNA
<213> artificial

<220>
<223> primer

<400> 4
gcgggatccg aaccggtcca gctacaggta 30

<210> 5
<211> 32
<212> DNA
<213> artificial

<220>
<223> primer

<400> 5
gcggaattcc tcgaggatgg gcaggtcttc ag 32

<210> 6

<211> 26
<212> DNA
<213> artificial

<220>
<223> primer

<400> 6
gcttcttccg caagatatac actccc 26

<210> 7
<211> 27
<212> DNA
<213> artificial

<220>
<223> primer

<400> 7
gaggatgcag caatagccac ataagac 27

<210> 8
<211> 24
<212> DNA
<213> artificial

<220>
<223> primer

<400> 8
tatcttccct ggttcctgaa cgac 24

<210> 9
<211> 22
<212> DNA
<213> artificial

<220>
<223> primer

<400> 9
ccgcagagct tctgtgtaat cc 22

<210> 10
<211> 25
<212> DNA
<213> artificial

<220>
<223> primer

<400> 10
cctccctatt ccccaagatg tatgc 25

<210> 11
<211> 22
<212> DNA
<213> artificial

<220>
<223> primer

<400> 11
ggagaggcta ttcggctatg ac 22

<210> 12
<211> 25
<212> DNA
<213> artificial

<220>
<223> primer

<400> 12
ctgggctccc tcttagaata accta 25

<210> 13
<211> 22
<212> DNA
<213> artificial

<220>
<223> primer

<400> 13
ggagaggcta ttcggctatg ac 22

<210> 14
<211> 21
<212> DNA
<213> artificial

<220>
<223> primer

<400> 14
ctctgagccc agaaagcgaa g 21

<210> 15
<211> 23
<212> DNA
<213> artificial

<220>
<223> primer

<400> 15
gaccagagcc tcttcatcaa tgt 23

<210> 16
<211> 22
<212> DNA
<213> artificial

<220>
<223> primer

<400> 16
gagaatctgc gcttacgagg ga 22

<210> 17
<211> 32
<212> DNA
<213> artificial

<220>
<223> primer

<400> 17
gcgaattcta tgtctggggg caaatacgta ga 32

<210> 18
<211> 37
<212> DNA
<213> artificial

<220>
<223> primer

<400> 18
gcggatcctt atatttcttt ctgcaagttg atgcgga 37

<210> 19
<211> 57
<212> DNA
<213> artificial

<220>
<223> synthetic sequence

<400> 19
ttggggtcat tgtcgggcat tggggtcatt gtcgggcatt ggggtcattg tcgggca 57

<210> 20
<211> 88029
<212> DNA
<213> Homo sapiens

<400> 20
gatcatgagg ttaggagttc gagaccagcc tggctgatat ggtgaaacgc cgtctctact 60

aaaaatacaa aaattagctg ggcgttgtgg caggtgcctg taatcctagc tacttgggag 120

gctgaggcag gagaattgtt tgaacccagg aggcggaggt tacagtgagc cgagatcacg 180

ccattgtact ccagcctggg cgacaagagt gaaactccca tctcaaaaaa aaaaaaaaaa 240

aaaaaaaaaa agacatgtat tctctctctc agtcacggac ggcagaagtc cgaagtgagg 300

agtgggcagg gctgcacttc ctaggctctc ggggagactt tttttcctgt cccttccagt 360

ttctggtggc tccaggcatg ccttggctta tggcagcatt attattccag tgtctgcctc 420

tgtgatcata gtgcctcctt ttctttttt ttttttttaca ttttttttt gtatttagag 480

aaaaaaacac ttaacataaa atttaccatc ttaacctttt ttttgagact ctgttgccca 540
ggctggaatg cagtgttaca atcacagctc actgcagcct caacctcctg ggctcgtgac 600
atcctcccat ctgactctcc caagtaactg gggaccactg gcatgtgcca ccacacttgg 660
ctaattttta catttttgt agagacaggg tttctctatg ttgcctaggc tggtctcaaa 720
ctcctcagct caagcaatcc tcctgccttg gcctcccaaa gtgctgggat tataggcgtg 780
agccaccacg cctggccatg ttaaccattt ttaggtgtgc agttcagtat gttaaatata 840
ttcacattgt tatgaaacag atgtccagaa ctttttcatt ttgctaatct gaaactctgt 900
acccattaga caacagctcc ccccgcaggt aaccattcta ctttttgctt ctatgatttt 960
gactacttta gacactttat ctaaatggaa tcatatggca attgtctttc tgtgattgac 1020
ttatactact tagcataatg ttaagtttca tccatgttgt agcatgaatc agaatttcat 1080
ccctttttat ggcttgataa tactgcattg tatgtatata ccacattttg cggtaggtac 1140
aatgtatatt tacattgctt ccacctcttg gctactgtga ataatgctgc tatgaaaatg 1200
ggcgtgtagg tatcttttcc agatcctgac tttacttcct ttggataaat acttacaggt 1260
gggactgctg gggtatatga ttgttctact tttaattatt taacactctt ctacaattta 1320
ttttctgttt ttgttgttct aatagtagtt attattaggt gaggtatttc ttatctctta 1380
taaggacacc tgtcattgga tttagggtcc acctggttaa tccaggatta tcaagtctca 1440
aaatcctgaa ttacatctgc aaagactctt tttccacata aggtcacatt cacaggttcc 1500
agtgattcaa acatggacat gtcttctggt cccccattat gtccactata ctctcttttt 1560
ttttttttt tttaagatgg agtctcgctg tgtcgcccag gctggagcgc agtggcgcga 1620
tcttggcttc ctgcaagccc acctcccagg ttcacgccat tctcctgcct cagcctcccg 1680
agtagctggg actacagaca cccgccacca cgcccagcta attttttttgt attttttag 1740
tagagacggg gtttcaccat gttagctagg atggtcttga tctcctgacc tcgtgatccc 1800
cctgcctcag cctcccaaag tgtcgggatt acaggcgtga gccactgcgc ctggcctaag 1860
tccactatac tttcttcttc cctgccttat ttttattctt gatacttatc tccatctgac 1920
atgctctata tttctttatt tatcttgttt ggcagacgac aatcaagata aagccatgga 1980
gacaaggatt tttgttgctg ttgttcttgt tttttgagac agagcctcac tctcacccag 2040
gcctagagtg cagtggcaca atctcggttc actgcaacct ctgcctccca ggctgaagtg 2100
atcctcccac ctcagcctcc agagaagctg ggactacagg tgcttgccac atgcctggct 2160
aatttttgt atttttggta gagacggggt tttgccatgt tgtccaggct agtcttgaac 2220
ttctgagttc agatgatcca cccaaagtgc tgggattaca tgcgtgagcc actgcgcctg 2280
gcctagacaa ggatttttgt tttggtcacc tgtgttttcc cattagaaca gtggctggca 2340
caaatggctg cacagcacat actggttgaa caaatgaagg acggggtggc tggtctagac 2400

aaagagccta gacaaacatc ggcagaaatt gcttcatggc ttctgagcag aaaaatctct 2460
catctgggga attagactcc ctaagttaaa ttttctttct tttttggaga tggggtctca 2520
ctctgttgcc caggctggag tgcagcagca ccatcacagc tcactgcagc ctcaacctcc 2580
tgggggtcaa gcaatcctcc cacctcagcc tcccgagtag ttgggactac aggcccatgc 2640
caccatgccc agctaatttt ttttttggta gagacagggt gtcaccatgt agcccagact 2700
ggtcttgaac ttctggactc aagcgatctt cctgcctcgg cctcccaaat gctgggatta 2760
caggcatgag ccacagtgcc tcacctccta agttaaattt tctgcagtgg agaatacaat 2820
ctctttaata ttatctctca gttaagacaa atttcaggat cctccttaaa aaaaaaaaaa 2880
aaaagaaaga aaataagttt gccaatacaa ataccatttc tcactaaagt gaattagggt 2940
tccttggaga aatggttggt tttgtttctg ggcagtaaat gtataaaacg gaaagcaagg 3000
aagtccaggg tgtccaatct tttggctttc ctgggccaca ctggaagaag aagaattgtc 3060
ttgggccaca cataaaatac actaacaata gctgatgagc taaaaaaaaa aaaatctcat 3120
aatgttttta gaaagtttat gaatttgtgt tgggctgcat gggccacagg ttggacaagc 3180
ttgacctaaa gactactagg attgtggatt actaggattg tgccagaagg acacagcagc 3240
aactaaatat ttgatgagac aatctgaaca tttaaaaaag gacaatgact gtaatggatt 3300
aaagcacatc aaatatctaa acatccatca attcatgata gcactgcccc ttctccccaa 3360
agaacccaaa gtggtcacag ttagaggttg ctggggcatc catccatcca tctttattat 3420
tattactatt atttgagaca aggtctcact cagtcaccca tgctagagtg ctgtcgtccc 3480
atcacggctc actgtatcct caacctcctg ggctccagcc atttccctgc ctcagcctcc 3540
taagtagcgg gaattacagg catgcatcac catgcctgtc taatttttac atattttgta 3600
aagatcttgc catttcctgg gctcaagcag tcctcctgtc ttggcccccc aaagtgctgg 3660
gattataggc agagccactg tgcgtgggaa agcatcaagc atacatcctg gctatcccga 3720
atggattgta tttcaaagta accagagaga tgagggaatg ctcacctttg tagaagaatc 3780
tcatcttata aatgcaggag aaatgagagc atttgaaatt accactttgc acacctaagg 3840
aacatcataa aactacacta gggtttctca acctgggtac tactgacatt ttgagctgga 3900
taattcttcg ctgtgggggg gaggtgtgct ctgtgaatta tacaatgttt agcagtattc 3960
cttgcattca ttttctagat aacagcagta ccacccgcca ccccccaccc agttgcaaca 4020
atccaaaata tctccagaca ttgccaaatt tccccttggt aggacagggc agaatcaacc 4080
ctcgttgaga accaatggtc taatgatcat caacgtttgc tagactatta gaagaaaggc 4140
tgatggtaaa cttcatggat aatcaggatg acaacccccа aattgagaga tgaattacaa 4200
tattactaag agacaaaccc gccttgtgcc tcagtagaag tacatagtgc cagccacgaa 4260

gcgttattgc aaacaaaaca aaacaaaaaa acccaaacct caacattaca cctaaaccta 4320
atgaagcttc tagccagggg caaatccaag ctttgtgggg ccttaaacta tacaaatttc 4380
acagtcctct ttaagaaaaa gacacaaaat tataaatgcg aaattaggta cgggggtcta 4440
tgcaagggag ggcctgaaga ttaagcttca ttagtttcac tgtaaacctc ccctgactct 4500
agaattaact gtgattacag gacataccag ggacaaaaaa cgttaaatga cacctgaaga 4560
tacaatcagc aaaacccaga aagtggaaaa ttctgttggt caaatgaccc agtttcttca 4620
ataagtaaat gccatgaata acaaacaaca aaaagagagg ggaaatttat atatataata 4680
tatatataat atgtataata tatatatatg ttgttatatg gtttgttttt ttttttttgg 4740
acacgagtct ctctctcacc caggctggag tgcagtggca tgatctcggc tcactgcaac 4800
ctctgcctcc tgggttcaag cgattcttct gcctcagtct cccaagtagc tgggactaca 4860
ggtgagcacc accacaccca gctaattttt gtatttttgg tagaggtagg gtttcaccat 4920
attggccagg ctggtctcga actcctgacc tcgtgatctg cccaccttgg cctcccaaag 4980
tgctgggatt acaggtgtga gccactgcgc ccggccctgt ttttgtttgt tttgagatag 5040
aatctcactc tgttgcctag gctggagtac agtggcatca tctcagccca ctgcaacctc 5100
cacctccctg attctagcaa ttctcctgcc tcagcctccc aagtagctga gattacaggt 5160
gtgcaccacc acacctggct aatttttgta tttttagtag aaacagggtt tcaccatgtt 5220
ggccaggctg gtctcaaact cctgacctca agtgatcctt ccacctcagc ctcccaaagt 5280
gctgggatta caggcatgag ccactgtgcc cagccaaaat tgttatatat taagagacat 5340
atatttgtat gaaatgcagt aagtaaacct tgtttggacc ctaaatatct aatgtacaaa 5400
atttttaag gcaatgggga aaattaaaca catactaggt attaagtgat gttaaataat 5460
ttttaaaatt ttggtgggtg tgataatagt ataaagtcct tatctgttag agacacacac 5520
tgacgtattt ataggtgaaa tgacatgatg tccaggattt gctttaatat acagcacttc 5580
aaaaaaaaat gcagaaaggg atacatgaaa tgagaaaggc agcaaactgt tgttgaagtt 5640
ggatgatgag tacagccctc cgttattatc caaggggat aattcctgga cccctacgga 5700
taccaaaatc caggtatgct caagttcttt atgaaagttc attgtaatta ttctattata 5760
taaaagtttc gaaattctgt tgataaaatg tatttttagt agagacgggg gtttcacctt 5820
gttggcccaa ctggtctcga atttctaacc tcagatgatc caccagcctt ggccgcccaa 5880
agtgttagca ttacaggcgt gagccaccgc gcccgcctgg ccttgataaa aacagtttta 5940
accttccgtt gcttcgattc catgcccact aagtaacatt ccagtttgtt tttcactttc 6000
aaaaggatgt gctgtaacta ggggatgtaa acaagctcca tgaccctact tattccaagt 6060
tttcgttcca ctctcccacc tttttttaa gacaaggtat caccctcggt cgcccaggct 6120
ggagcgcgat cactgcttaa tgcatcctcg acctcctggg ttcaagcgat tctcctgcct 6180

cagcctccca agtagctggg actacatgcg cacaccacca caccggttaa ttttttgtag 6240
agacgggggt ttcaccatgt tgcccaggct ggtcccgaac tcctgggctc aagggatccg 6300
cccgcctcag cctcccagag tgctgggatt acaggtgcca gccaccgcgc ccggccccag 6360
cttcttaaaa gaatgatccg aaactatggc agcactgggc ttttggtccc cacccaagaa 6420
atgcccgctc gcagaggctc gccgcggcag gctctcccga cgtgacagag tgtgggtctg 6480
gattcagcct cggttcttac gagtcagata ggtggacacg caaagcaaaa catcacaggg 6540
ctttttgtat ttagcacaga aaacacttgt gagcccgagc tgagaaccca aaaggcacgc 6600
ttcaggccat cgtagccacc aagcctggtc agattccgtc caccgtctcc ttggtgctcc 6660
gagacccaaa tcgctgactg gggccgaggg cgggcgtgac tgcgcaggcg tgcctccccc 6720
gcgagatgcc ggaggtaagc tgcggggtaa ggggcgagaa attaagggcg aacgtcattg 6780
cgcatgcgcc ctctactctc gttgcggggg taggcgggcg ccgggctgtg tgagggggcg 6840
gggcgcggca gtgttcggta cggatggagt tgcaggagac ggcgagtaca tatcactgcg 6900
caggcgtcct cttcccctaa ctctcagggt cgctagggtg gcgcgcaggc gcagagcgat 6960
gcgcaaatgt gcgcaggcgc ttaggggctg aggcgcgatg gcaggtgtcg gggctgggcc 7020
tctgcgggcg atggggcggc aggccctgct gcttctcgcg ctgtgcgcca caggcgccca 7080
ggggctctac ttccacatcg gcgagaccga gaagcgctgt ttcatcgagg aaatccccga 7140
cgagaccatg gtcatcggtc aggcgggctg agggtgggga ggccctttgt acccagctca 7200
gccctcggcg gcgctccctc ctccgagcc cagccgggtc gctggctccc ccagtaccta 7260
gcctgagggt gccccgagga cgccaggccc cctgcctaga gctccgggcc gcacgtcgga 7320
gggggccggg cggagaggcg gcccactagg gccggtcgtg actatgtgtc tgccccgcag 7380
gcaactatcg tacccagatg tgggataagc agaaggaggt cttcctgccc tcgacccctg 7440
gcctgggcat gcacgtggaa gtgaaggacc ccgacggcaa ggtaaggctg gcgttggccc 7500
acgcagccgt tcttcagtgg agctcccgtg gggtgtaaag cactgcctgg aggaggcctc 7560
aagggacagg aacttgcact tggagagcct gcggtataaa ggtggggcct tcactcacat 7620
atgttgcagg tggtgctgtc ccggcagtac ggctcggagg gccgcttcac gttcacctcc 7680
cacacgcccg gtgaccatca aatctgtctg cactccaatt ctaccaggat ggctctcttc 7740
gctggtggca aactggtaag aggattttct ctttggcttc agcttagaat ctctcacttg 7800
tttccaaatt ttgatttatc aagattgtga aactttgtag cacagtcaga attggggaga 7860
cagatgttgc cttctgctcc acagccaggg acaatagtgg gttccatacc ctggaacaga 7920
caactggagg ccccaccact catacattcc atgtttcctt gtagcgggtg catctcgaca 7980
tccaggttgg ggagcatgcc aacaactacc ctgagattgc tgcaaaagat aagctgacgg 8040

agctacagct ccgcgcccgc cagttgcttg atcaggtgga acagattcag aaggagcagg 8100
attaccaaag ggcaagtgca tatctccttg taatttgaga gggcagttga cctttatacc 8160
cactatacct actcaagttt ctgcttggga gatcagctct gcagagaatg gaatgagaag 8220
tattggttta gataggttgt ttgtttgttg tttttgagac ggagtttcac tcttgttgcc 8280
catgctggag tgcaatgcca tgatcttggc tcactgcaac ctccgcctcc ccaggttcaa 8340
gcgattctcc tgcctcagcc tcctgagtag ctgggattac aggcatgcgc caccatgcct 8400
ggctaatttt gtacttttag tagagacggg ggtttctcca cgttggtcag gctggtctcg 8460
aactcccgac atcaagtgat ccgcccgcct cagcctccca aagtgctggg attacaggtg 8520
tgagctaccg cgccctgcct gttttgcttt tttatcaaaa cattttattg tggtaaaata 8580
taacaccaaa tgtgtcattt taactgtcta tatagttcag tggtattaag tgccttcata 8640
atgttgtgct accaacacca tcatccagct ccagaacttt ttcatcttct caaactaaaa 8700
atctgtactt attttgtttt gtttttgaga tggagtctcg ctctgttgcc caggctggag 8760
cgcagtggcg ccatctcggc tcactgcacc ctccgcctcc caggttcaag cgattctcct 8820
gcctcagcct cccaagtagc tgggattaca ggcaagtgcc accatgcgtg gctcatttt 8880
gtgtttttag tagagactgg gtttcaccat gttggccagg ctggtcttga actcctggcc 8940
tcaggcaatc cactgccgca gcctcccaaa gtgttgggat tacaggcgtg agccactgca 9000
cccagcaaat ctgtacttat tataaacaat aacttcccgt ttccttttgt cctgacaccc 9060
accattctac tttctgtctc tatgatcctg actaccctat ctcatataag tggaatcatt 9120
cagtatttgt ccttttgtga ctggcttatt tcactgagta taatgttctc acagttcatc 9180
catgttatag catgtgtcag aatttcttaa ggctaatatt ccattgtatg catgtgccac 9240
atttcgcttt cagtagtcat ttttaagctc tataaaataa aatgaagaaa ggacagttca 9300
caatctagta atagccattg cctacctgtt tttcttggac tcttgttgga aatggtagga 9360
tcatgatttc agtcctaaca gagatgcttg tggagggaca gcctgtccct ttcttggggc 9420
agcctcagtg gggagaccat agcactccta atggagtcac agatagtatt ccaaaaggag 9480
tttggtcctg gagttgagta attacacgca gggagggacc tcacaacagc cagactgttt 9540
ctcctgctca cttaaccctg tgttgcccca cacagtatcg tgaagagcgc ttccgactga 9600
cgagcgagag caccaaccag agggtcctat ggtggtccat tgctcagact gtcatcctca 9660
tcctcactgg catctggcag atgcgtcacc tcaagagctt ctttgaggcc aagaagctgg 9720
tgtagtgccc tctttgtatg acccttcctt tttacctcat ttatttggta ctttccccac 9780
acagtccttt atccacctgg atttttaggg aaaaaaatga aaaagaataa gtcacattgg 9840
ttccatggcc acaaaccatt cagatcagcc acttgctgac cctggttctt aaggacacat 9900
gacattagtc caatctttca aaatcttgtc ttagggcttg tgaggaatca gaactaaccc 9960

aggactcagt cctgcttctt ttgcctcgag tgattttcct ctgtttttca ctaaataagc 10020
aaatgaaaac tctctccatt accttctgct ttctctttgt ccacttacgc agtaggtgac 10080
tggcatgtgc cacagagcag gccctgcctc actgtctgct ggtcagttct gggttcactt 10140
aatggctttg tgaatgtaaa taaggggcag gtcttggccc tagaggattg agatgttttt 10200
ctaaatctta gaactatttt tggataaatt atatattttc cttcctagta gaagtgttac 10260
tgcctgtaac tagctcaaaa taccaatgca gtttctgcat tctgggtttt gtttttcctt 10320
ttttttttt ttttttttt ttgagttttg ctcttgtcgc ccaggctgga gtgcaatggc 10380
gtgatctcag ctcactggca acatctgcct cccgggttca aatgattctc ctgcctcagt 10440
ctcctgagta gctgggatta caggtgcccg ccaccacgct cagctaattt ttgtattttt 10500
agtagagatg gggttttacc atgttggcca ggctggtctt agactcctga cctcagttga 10560
tccacctgcc tcagcctctg cattcagttt attcacatat ttttggtaac tcccatggca 10620
gctcctagga tttcagcggt ctgtgggcca gaaagcaggc accagggctg acctcaaggc 10680
cgtatcagag ggccaagcag agttcttttg gatacctgct tttcatccca cagggcctta 10740
gagtcagagg taaggtagca acagagctag aatggggcaa tgcactctta ccctccttct 10800
caacttttat ttaagctgtg ctaaatgttt tcttcaaggg aaccagattt agttctttac 10860
agaattttcc agtgaaataa aacatgttgt aatagctgtg tttgagatga aataagaggt 10920
tgtgggtaga ggggaggcac ctaaaggaaa agaggaaagg tgcctgggct acctatgcag 10980
ataacctgga gtggacttca ctgtggactc gtggtactaa ggcttggcct ggacaggcag 11040
tctagggggt atgggaatac acggtgtggt tgttcaacta tttgcaaagg tcaaccaaat 11100
agaccacatg ttcgcaaagt atcatctgag gaaattaagt accttcttag ccctctcagt 11160
cataaatttg aacaaatttt aatacacttc cctcatgccc ttctatataa aacttaatac 11220
cattagttcc ccattcttga cattttattt cagtttttat tatatattta tttgaaatat 11280
ttattaaatt atctgaccta cagaactaaa ttcttctcct tttgttattt cttatgtcct 11340
ataccatata tgtacctatt tatatatata tttatgtatt tttaaaattt ttatttattt 11400
tatttttga gacagtcttg ctctgtcgcc caggctggag tgcagtggca tgatcttggc 11460
tcactgcaac ctctgcctcc cgggttcaag cagttctgcc tcaacttctg agtagctggg 11520
attacagaca cccaccacta cacccggcta atttttgtat ttttatttta tcttattcat 11580
ttatttattt ttgagatgga gtctcactct gtcgcccagg ctggagtgca gtggtgcaat 11640
cttggctcac tgcaacctcc acctcctgag ttcaagagat tctcctgcct cagactcccg 11700
agtagctggg attataggcg cccgccacca tgcccagcta atttttgtat ttttagtaga 11760
gacagggttt caccatgttg accaggctgg tcttgaattc ctgaccgcag gtgacccgtc 11820

tcgcctccca aagtgctggg attacaggtg tgagctggcc gggcacaggt gatggggtct 11880
tgctctgtcc ccaaggctgg agtgcagtgg tgccatcaca gctcaaagca accttgagct 11940
cccaggttca agtgatcctc ctaccttacc ctcccaagta gctggtacta caggtatact 12000
ccactgtgcc tggctatttt tactcttaaa aatacatgtg ggctgggcac ggtggctcac 12060
gcctgtaatc ctagcacttt gggaagccaa ggtgggtgga tccctatagc ccaggagttc 12120
gagaccagcc tgggcaacat ggcgaaatct tgtctctgca aaaaatacaa aaaatttagc 12180
tggtggcaca tgcctatagt cccagctact tgagaggctg aggtggaaag atcacttgag 12240
cccgggaggt caaggctgcg gtgagccatg atcgtgccac tgcactccag tctgggcaac 12300
agtgatccca tctgaaaaaa aaaaacaaaa aaaaaaatgc aatttagggc caggtggggt 12360
ggctcacgcc tataatccca gcactttggg aggccaaggc aggggatcg cctgaggtca 12420
gcagtttgag accaggctgg ccaacatggt gaaaacccct ctctactaaa agtataaaaa 12480
ttagccaggc atggtagtgt gtgcctgtaa tcccagctat tcaggaagct aaggcaggag 12540
aatcgcttga acccgggagg aggttgcagt gagcagaaat cgagccactg cactccagcc 12600
tgggggcag agggagactc tgtctcagaa aaaaaaaaaa aaatgcaatt tagttctcta 12660
ggcttttcca tttaatagtt ttatatcctc ctgtttctaa atctggatga cagtgtaaca 12720
ctccagtaag gtgaattgtg aattgctgaa attcttcaga tgtttaaaag agttttcagt 12780
attcctcatg ttagaattaa tgcagagaaa aattttatcc tttgaactag ttacatgttg 12840
tggacttctg gcctgaggct cttggggatt atgtgacata ttgggaaggg acacatttct 12900
gctctgtggc tgttactaga aatctagcca gcaaatcaga ctacgtttgt gagaagacag 12960
gaaggcacag attagggttg agccagcctt caacaggttt ggctggcagt agacacagtg 13020
gagcacatct taactatttt ggtaggtcct gggtttctct tggtagtttt tgatagaaag 13080
gggaatggtg tgaggaaaaa gtgggcatac atttcacctt tccactgata aggcaggtgg 13140
aattgggata gtcagtggat gggccaatag ctggtggctg tgagaagaat aaggatttcc 13200
atactggtgt gtcatattta cagataggtt gtgacctaaa aagtttttta aaaaacagca 13260
gttagggcct gggcgcggtg gctcacgcct gtaatcccag cactttggga ggccgaggcg 13320
ggcggatcac aaggtcagga gatcgagacc atcctggcta acaacggtga aaccctgtct 13380
ctactaaaaa tacaaaaaat tagccgggcg tggtggcagg tgcctgtagt cccagctact 13440
cgggaggctg aggcaagaga atggcgtgaa ctcgggaggt ggagcttgca gtgagctgag 13500
atcatgccac tgcactccag cctgggcgac agagtgagac tccttctaaa aaacaaaaac 13560
aaaaccaaaa cagtagttag ggtacacaca cacaaattct agtgattttc cccccaatac 13620
taccccttgac ttttgaaatt cttgctttct cagagtttac aacatcctta ccaaacagcc 13680
ttctccctcc ttaccacaaa aaaagaaaaa aaagttctgg ggttgagggg acactccatt 13740

cttaacatcc tctattatcc cagcccaatt ccccagctct cactgggact agttgtacct 13800
atcttcatca tttggtccca gcatgactac ctgttggtgc atgagctgat ctctcctaac 13860
ctaacagcca gatgctagtc tctggtactc agatgctggg ctgcatcaga taggatgcac 13920
aggatcatcc tggaagcttg ttgacataga ttcctgtgca acactcagat atagtcttaa 13980
tgtagatttg tgttgggtgg tatggtaggt agaataatgg cctaccactc tgaaacatat 14040
gaatatgtta cctaacatga cagaagagaa ttaagttgct aatcagatga ctgtaaaata 14100
aattatcctg gatcatctgg atgggcctaa tgtaatcaca aaggttgttt ccttgccttt 14160
tccagcttgg ctctggctcc cttcctccag caagggtggg ttgagctctc acatggcacc 14220
actttgacct cttctgcttc cctcttctac actgaaagac ttatgggcca ggagcagtgg 14280
ctcgcacctg caatcccagc actttgggag gccgaggaag gcagatcgct tggccccagg 14340
agttcaaaac cgtcctgggc aacgtggcga aaccccatct agaaagaaaa gaagagaagg 14400
ggaggggga ggggaggagg agttacatat atacacatac acacacacac acacgtacgt 14460
acatacatac atacacgcta actggacgtg gtggtgcgtg cctgtagtct tagctttcca 14520
ggagactgag gtgggaggac cacttgagcc tgagatcgcg ccagcctggg tgacagtcag 14580
accatgtctc aaaaaaaaaa aaagatttgt gattaggatt cttagtcctc acctgtatta 14640
ttttcctatt gctactgtaa caaattacca caaatttact ggcttaaaac gacgcaagtc 14700
tgtaggtcag aagtctgaca cgggtcttaa ctggtgaccc gagtcagatt tgggacacaa 14760
agaacagaaa ccaagctgtg caggtttctg acaggcagtc cggttaggga gccctacagc 14820
aacccgccgg tcctctctct caggcagttg ctgccatggc tcattattcc aaccggttct 14880
cctcagccca gtctatctca gtggctccat tcatagggtg atgtgcccgg cgggacacta 14940
accctaacca agcagagaga cggtcatgcc cgtcacgacc tcggccctcg ccccggccga 15000
ggcttctcct gcaggtcgcg agaatcaggt gcgtcagcgg cgtccgggaa cgccggaaga 15060
gccagtggag cggctctgta gtccaaagta ccccgtcgac cccagcacgg ccgctccacc 15120
gcctcctact agacccagtc ctagggactg cgcagtcgca gagctccgtc cgagtaccgg 15180
aagcctaggc cgccagcact tccgggaagt gacttcgtct ccgaagccga ttggttgttg 15240
ctttgctccc gctcgcgtcg gtggcgtttt tcctgcagcg cgtgcgtgct gcgctactga 15300
gcagcgccat ggaggactct gaagcactgg gcttcgaaca catgggcctc gatccccggc 15360
tccttcaggt acacgcgagg gctggggagc cggcttacgg gctctgcggg gcgcgccatc 15420
gctcttcacg ccgcttaaac cgcactcctg gtctcctagg ctgtcaccga tctgggctgg 15480
tcgcgaccta cgctgatcca ggagaaggcc atcccactgg ccctagaagg gaaggacctc 15540
ctggctcggg cccgcacggg ctccgggaag acggccgctt atgctattcc gatgctgcag 15600

ctgttgctcc ataggaaggc ggtgggtaac gagagagctg aggggaggaa ggaggcaagc 15660
tccaaaagcc tgggaagggc ggttcccgtt tgtctgaggt tttctcttgg ccctgtaccc 15720
gtgcaggccg gcctgagaac ctggtgctgt tgtggcaaac actctgggct ggagttcagg 15780
ttacctggat ccttgtccgg ccctgctacc accaaccttt gcgtaatctt cgacaaagca 15840
ctttcttttc tttcttacat aaaaagggag cacatctatc ttttctactt acagaattat 15900
tgtgagaatt tagcttcata actagtatat ttaaagtagc ttcataaaca tcagagtacg 15960
ttattctttt tgagggtcag tgcctgggga aagaactctc cactctgcat tctgaggcgg 16020
gcagagtgat agatgatcaa agtactgcta agtagtgttg cagcagatgg gtcaggtagg 16080
ctggaagggg tagagacacg tggacacagt gatgtgcact gctggctaaa gtctttaatt 16140
catattctta cagacaggtc cggtggtaga acaggcagtg agaggccttg ttcttgttcc 16200
taccaaggag ctggcacggc aagcacagtc catgattcag cagctggcta cctactgtgc 16260
tcgggatgtc cgagtggcca atgtctcagc tgctgaagac tcagtctctc agaggtgggt 16320
aaaagcagca aagctgtacc tgaatgaagc tacacagtgt tgtggggttg ggtttgtgtg 16380
tggcaaaaaa gagagcaaat ccagggtgag atcccagctg ctacattctg cctgatactg 16440
atgtcttgtc cacctccaga gctgtgctga tggagaagcc agatgtggta gtagggaccc 16500
catctcgcat attaagccac ttgcagcaag acagcctgaa acttcgtgac tccctggagc 16560
ttttggtggt ggacgaagct gaccttcttt tttcctttgg ctttgaagaa gagctcaaga 16620
gtctcctctg gtaaggcaga ggtgggtgtg attcctagtg gaaacatctg tgagtaggag 16680
ttgggacgag agcggggtgg ctggaagcca gttactacaa ttagcggccc ttggagctgg 16740
aatctgattg gattctttca tttcagtcac ttgccccgga tttaccaggc ttttctcatg 16800
tcagctactt ttaacgagga cgtacaagca ctcaaggagc tgatattaca taacccggta 16860
agaggcacca tggaagtgtc tggagctgca gacatggggg cactcaaaga tcttgatgct 16920
ccttcttagg ggattctttg gtgttttggg tgggacagtt gtcacttagt gtctcatccc 16980
tggtcctgag gcactaaaag ccagtggtct aaaatcacta tatatttcca agtgtccaca 17040
agggatgtct cccatttcag gccatgcttt gcctaaaatc ctgagcaagg acctccccta 17100
aggggcagct ttgagcagca gagccaaaat tctaaggcca aggttctcat cttaagtaaa 17160
ctttaccttt cagaaggcct gttgctgtag gccttccctt ctcaatgtag tcctttattg 17220
atgtgtttct ctttgttctg tgcttggaag tattttatat atggtttata tggtatactc 17280
tatataccac aacaataagg gcattttggg gttttaggtt acaaaactgg aggagagtta 17340
gggtgccagg aatccttaaa tgcatctctg ccctgcacta aaatgttgat gctttggttg 17400
gtgagtaagt ggccatacat ctctgtgttc ttttcctttc tgaccacagg cctgttttct 17460
cccccaggtt acccttaagt tacaggagtc ccagctgcct gggccagacc agttacagca 17520

gtttcaggtg gtctgtgaga ctgaggaaga caaattcctc ctgctgtatg ccctgctcaa 17580
gctgtcattg attcggggca agtctctgct ctttgtcaac actctagaac ggagttaccg 17640
gctacgcctg ttcttggaac agttcagcat ccccacctgt gtgctcaatg gagagcttcc 17700
actgcgctcc aggtctgcca cagccaacat cttggttgaa ataagttgaa gatagagatg 17760
gaaaggggac ccagttaatg ttctgtttct taagcactta gtaggggcca ggttctagat 17820
gtgactgata ctgacttctc ccaactccaa aatacctatc atggccgggc accatggctt 17880
atgcctgctg taatctcagc actttgggag gccgaggtgg gcggatcgcc tgaggtcggg 17940
agttcaagac cagcctggcc agcatggtga aaccccgtct ctactaaaaa tacaaaaatt 18000
agctggacat ggtggcaggc acctgtaatc ccagctactc aggaagctga gataggagaa 18060
ttgcttgagc ccgggaggtg gaggttgcag tgagccaaga tcgtgccatt gcactccagc 18120
ctgggcaaca ggagtgaaac tctgtctcaa aaaaacaaaa ccctataatt atttccagct 18180
gaggaaactg aggcacaatg attaagtagg gaaagagatt aagaagagga aaaaggaaag 18240
ggtgatggtt actgtgatac tagggatggc agaggggcct tgagcttgct ctgctgagct 18300
gattctctgt ccgctcttgg ctgcaggtgc cacatcatct cacagttcaa ccaaggcttc 18360
tacgactgtg tcatagcaac tgatgctgaa gtcctggggg ccccagtcaa gggcaagcgt 18420
cggggccgag ggcccaaagg ggacaagtga gtccatgcct ctttttccat ccctccccag 18480
aaatgcctgt gtttttagct ttttggaaga ctaaaaccag agtgcacaga gcagggagcc 18540
aaaccttcca ggcctggctg gtagtgtagc ccagagagcc ccacaggttc ttgctcagct 18600
gcctggatat agagaaggga gtggatggtg cacactgcac atgcaccacg aagggcaaaa 18660
ctgccggggt tgttggcatg cagagccctg caggggagat ggcccatcct gcattggtgg 18720
tatggctgtg acttgcaggg agcatatttc tgaagggaaa aggaaccccc caactctcca 18780
gtctctgtcc agctgaaggc ttgactagct cagagttggt tttcagatca ccatgtaggg 18840
caatgagttc tgctgttgtc ccagaacaga ggtcaggccg agatttgggt acatgtcaaa 18900
gctccaggct gccccaggaa accctgactc ctggaacggt tccattgttg gagagtcctc 18960
tgtatgtcag ggtcttatga tctacaggca tttagaggaa gttttgctga ttcagcgtgt 19020
gaatacgtgc ccagaggaga ggaagggtcc ggctgacatt gagttatctc tgcagggcct 19080
ctgatccgga agcaggtgtg gcccggggca tagacttcca ccatgtgtct gctgtgctca 19140
actttgatct tcccccaacc cctgaggcct acatccatcg agctggcagg tagtagtgtg 19200
acggcccagg catctgcatg gtaggcacac tgagggactt ggggtgtgct ggacagagcc 19260
tgcgggttgg agatgcaagc tgcactgtct tcccttgcag gacagcacgc gctaacaacc 19320
caggcatagt cttaaccttt gtgcttccca cggagcagtt ccacttaggc aagattgagg 19380

agcttctcag tggaggtaag agcctggctc ttgtggtcct gggccagggt caggcttctt 19440
ccacaatgct ttaaaactcc atgataatga tgacagaggt cacaacatag tgtgacaggc 19500
cacttccacc atccatcctt gttctgccct gagtggcagg cactgtcccc cttgagagat 19560
aaacaaattg aggtaatttg tccaaagttg tgtttactgt ctgcctcatg agcgttgagt 19620
gacctgacag gctgctgtga cagctcagga cagcacctga ccccagggtg ctgggtggtc 19680
ctggactgct ctctgtggcc gtcgtcatgg gggtaccttg actcccaagg aataccatgg 19740
ggtactcctt gggagaggag aagagagtgg gtgacgggtt cttgggcttg gggccacaca 19800
ggccaccccc atccacacac ggggacagat gggtcatcac tgtaagaggc ccaggtgcag 19860
ctaacctgca tgttcggcat cccaggaagg cggtgggtcc cctgctgctt tcccccaagg 19920
gggaggtgca ggaggcctcc aatgaagacc ctatcctaag gcctcagcct gtgggaccct 19980
cgctgctttc ttctccacag agaacagggg ccccattctg ctcccctacc agttccggat 20040
ggaggagatc gagggcttcc gctatcgctg cagggtgagc tgctgtggtg gggaggggaa 20100
tgagagggga ggggctgtgg cccagggatt gcaccgtctt gctgagcatc caggtgtgaa 20160
gggaggattt ggggcagcct cactgtcttg accttcagtg tccaccccca ggatgccatg 20220
cgctcagtga ctaagcaggc cattcgggag gcaagattga aggagatcaa ggaagagctt 20280
ctgcattctg agaagcttaa ggtgagtgga tgggaggtga gaaggggata gatcttagac 20340
ggctgccctt tttggagact ggctgagctc cgagtggtga gaagcagaga actgggcagt 20400
tttctggcct ttggcacgga aggggaggaa atggacccag aatcatggaa ggaagccagt 20460
ctgttctgct tggtggtaaa ttggcacaac cttatggtgg acactgtcca gcagaattac 20520
gagctcatgt gtcctttcat ccgaaattcc acttctggaa cttaatcctg gtcacgcttg 20580
tgaatgtgca cagtcaagca tgtgcctgca ttcatccatc catggcatta tcatggaacc 20640
aaaagatgga aacagcctgg ggccaccata gggggcttgc taggtaaact caggtgcatt 20700
cagagccgaa ggttacatgg gaaggaatga ggttggttgc gtgtccatat ggaacagtct 20760
gtaagatgat gcccagcaaa aaggggtaca gggtactgcc atgtgtgtca tggagaaggg 20820
aaaatggaaa catccactcc cgggaggttc tgagaaatgc acagaagcag ctgcctcatg 20880
ccttttgaaa cacatgagtg tgttatcctt tgaaaagcta ggtctgtgaa gtcacagaag 20940
aaagatgctc actctgtggc tctccctctt cccccggcag acatactttg aagacaaccc 21000
tagggacctc cagctgctgc ggcatgacct acctttgcac cccgcagtgg tgaagcccca 21060
cctgggccat gttcctgact acctgggtga gtgtggcctg acagggcagg aggcagcagg 21120
ctggggaagt ggcattaatt tctccactgc tgggtcagcc cctgtgcttg gtgctgggga 21180
tgctcaggca gaatagaacc tggagaccct ggcagcacgc gggcatgtaa acaggcacac 21240
ccctgtgttt ctaaacttgt ttgcttggtc ccacgggtta gctgttgctg tctccatttt 21300

agagatgagg aaattgaggt agtgcagggt gggtggcaga cccagcattt caggccaggt 21360
cgtctccaga gctgggccaa atggccatcc atgggtcgaa gggagtgaac aggtttggga 21420
gagagtcacg ggcaggaggc agagagagcc acctgtgctg caaaagactc aagattagca 21480
gctgctgaag aggcatctgt ggagtctctg ggtaagaaca gtcagcaggg agacagactc 21540
tgtaaggcct aaaccgacaa ggtagcaaga gaagagccag tggtggtgcg agggatgcag 21600
gggttggcag gcatgaggtc aggaccctgg gattggtttc tgtagtgcag cccaggctag 21660
agctttatgt ggccattaat actgggcacc tctcctcatc tttggcaggc tctgggtaat 21720
gacttctttc agtgtctcat aggaggtgtt tttggtaatg agtatgtgtg acttttatgc 21780
ctaaaatgga ttgaaggagg agagtggtgg agaggaggct gtgggcagca agtgcaggac 21840
ccttcccaat gccacagggt ctgctcagcc tggacctgca gccacccagc gggtgtggtg 21900
ttgctgctat ggaggtgaca aagggtggag atggaatgtt ccagggcagg aaaagcctgg 21960
gcactgggaa aggaaggatc cagaagagat gggaacatga aaatgccaga gagagcggtg 22020
ggggccgggt tcccatggga cagtgagctg gaggagaccc cagtccaggt cctggcctga 22080
gatgtgagga ggggagttgg gagggtgggt agggagggag aaggtaaggc tagaactttg 22140
gcctcaggaa cccagtctgc tcgtatagcg gagtcatttg ccaaggtgtg gccaggaggt 22200
ttagaagggc caggagaagg tggaaaggtg tcaggatgtg ggatgtttga catttgaagg 22260
ggagggccca ggtgtggttg gcctggggga gtccatgggg tgggcgaggt gaagatagag 22320
ccaagatcag gtgcagctgg gatgcggggc cccctgtatc ggtagtaatg ggccacaggt 22380
gaagaaacta cctgttgact tttatttcag ctgcattttc tttctttaag gatgtctgtc 22440
tttttctttc ttgttacatg tttgttgtaa caaatctaaa caatatagga gagtgattta 22500
aatagtggaa gtctaaggtg ctcacattct cctggccctg tgcagatgtg gtagtgaata 22560
gatgtatgtc ataggctgcc agttgggtca gaattggaga atttgctgca gaatcagcgg 22620
gagggcaggg atgggagcag tagcggtgag cccactgctc aggcaagcat ctcttccagt 22680
tcctcctgct ctccgtggcc tggtgcgccc tcacaagaag cggaagaagc tgtcttcctc 22740
ttgtaggaag gccaaggtac ggctcctggg gactgcggac agccccagga ctcctcccaa 22800
cctgctcttt tgtcatcacc agaatgtgga ggcgccttgc cctagggagg ggaagagagg 22860
gtgccctagg gaggggaaga gggggcaccc tagaaccggg ccccaaaaat ctggtgtggg 22920
ataggggtac ttttgcagcc gcctgcaggc cctgcttttc tttccccagc tgcctttccc 22980
catttcctta tctgcagcac cttctggtcg tgttggccag ttgccggcac ggctcccttt 23040
gtgtctttct cagttgggtg ggtgggtggg tggattgtct gtcggcctga ttcccccaac 23100
taacctgtga ctttgcctcc ttagagagca aagtcccaga acccactgcg cagcttcaag 23160

cacaaaggaa agaaattcag acccacagcc aagccctcct gaggttgttg ggcctctctg 23220
gagctgagca cattgtggag cacaggctta cacccttcgt ggacaggcga ggctctggtg 23280
cttactgcac agcctgaaca gacagttctg gggccggcag tgctgggccc tttagctcct 23340
tggcacttcc aagctggcat cttgcccctt gacaacagaa taaaaatttt agctgcccca 23400
gtttgtgcct ccagcatatg aaaaggacta tttgaatccc caaaacatca ggagtcggga 23460
aacttcggaa gacagctgtg cctggctctg tggctgcatg cagtgcttca cttggccagc 23520
agaggtcagc tgtgccgagc tgccccagcc atgagaagag aagcctgccc ttgctggcag 23580
gtggctatgg ccggcccaga gccttcctgc ccagctcctg cagccctgct gcctgggatc 23640
aggctgggag atgggccttc ctgaccgcca gccttcctct ccccgagcac acgcacatgt 23700
agattcgggg ggaagctgcc tgctcttcct tagaggagcc ggggcagcta tctgctggtc 23760
cctttctgaa caactgttga tgtgtgagct gtgtctgtgt gttatgtgca taagcggtgg 23820
tgtgacatac acacatgtgt actgtccctt atgccctggc ctgagctctc cagctgcctt 23880
ctcagcctga aggctgggct tctctgctgg cttggggtcc tagattgcat gtcacctgct 23940
taccaggcgt cacaaggcca tgctgggggc atgaggaggt tggggcagca ggagagtggg 24000
gagaaactag gagagtgcct gagtatttta gaaagaacca agttttttct cggcaaaagc 24060
ttatacagag acgaaggagt ctgtgtcttt ggtcatggta ggactgaagc tagcaggacc 24120
cgagatttgg ggcctccatg atccctgctc ctcttctgtt aacacccaag gatttccacg 24180
aagccagtgt gtatgatggg ggcaggacag tggtactttc tgggcaggtg tgaactagag 24240
ctgctaagga gctgcagacg atattcttgc agtttggtgg ttagcagtat tcagaaggac 24300
aaagagttaa tggaactgga gataaagagc aaccatttga gcatctgctg ggagacatct 24360
gtcaactgca cagaccctat cagtgggcat cgctgccacc tcttggaaga caagacaggg 24420
cagagagtgc ctgcagtgct gaggcctggt ccttgcccta ggttggcctt cccacctggc 24480
ttcatggagt gctgaggctg gtcctgggga cagtgagtgc tctggatgtt ttagccaagc 24540
tgtgttctaa agtgatgcac agtctgtctc cactatgttt atctctctga ccctgtcact 24600
tccaagcaca ccctaccaag agcttgtatc ccagagccac cctgatggag aggagattgg 24660
tttccccagt gatttccttc tttgggggggt ggggtagagg aacatggagc cagccttatg 24720
ctgtattcgt gcctggggat agcagggtct gggcccggag cagaggagct tgggtaaaga 24780
tatggaggct gtttgttcaa agtgtacatt ccttcctcct aacggcatcc ctgggggaag 24840
ctatttctat gttttaggtg gggaagatga ggcttagaag ttgcctggtg aatgaggttc 24900
tttgcaagat ttgggttctg gcctgtccac cctggtggag tagctggtac cacgggggct 24960
tttgctgtgg ggttaggcac catgtgggcg ctctggggcc agggcattgg aaagaatggg 25020
aggattgctt gagcccagaa gttcgaggct gtatgatcac gcaactgcac tccatcctgg 25080

gcaacatact gagacactct ctctcttttt ttttgagaca gagtctcact ttgttgccca 25140
ggctggagtg tggtggtgca atctcagctc actgcaacct ctgcctcctg ggttcaagca 25200
attcttcccg cctcaacctc ctgagtagct gggattacag gtggccgcca ccacgcctgg 25260
ctaagttttt tatattttta gtagagacag agtttcacca tgttggtcag gctggtcttg 25320
aacttctgac ctcaggtgat ccacccacct cggcctccca aagtgctggg attacaggcg 25380
tgagccacca tgcctggccg tgagactcta tctttaaaaa ataaagaaca ggaaggtcca 25440
tcttcgtgtc ctgagactac agagagaaag taagtataaa tggctcgttc aacaccccac 25500
ctgggaggca ggtaccatgt gcccatttac gtgtgaacaa acaggcactc agggttggcc 25560
tcttggactt agtctggcca aagcctgtgc cctttgcaca aatgtgcaaa tcaggactgg 25620
ggcaggcctt ggatgagggt atgtgtgcta tgggcaaatg aacctagggc tgtccagggc 25680
caaacagcac agagggcatg tgggcctgga agggaggaag gaggtgtggc acatgctgcg 25740
tggaagccta aggcttcact aaacagcaga gaagcttgga tggttttcag gctggtgacg 25800
ccctgggctg aagcaggaag gtcaggagaa tgcagtggcc tctccactct gggctggcac 25860
agttttgccc acatgtatac ctgaatgggt gcctggctgt gtggactgtg ctatggtctg 25920
gaatcagatg gacaaggcac agtctatgag gcaaggagca gagatggtca gccaatgcag 25980
actgctcaat agtcatgttg ggagttcagg gtactggagg gctataaggg gccctcaccc 26040
agttggagag aatgctgcct tccttgagaa agtgaggttt atgctgagat gggaagggtg 26100
ggagggaaca gcaatcctag caggggagac agcatgtgca aattccctgg ggtgggaggg 26160
atccctgcac atttgagggt gaaaagacca gagggttgtt accaaaatat gcaatggggg 26220
ctgaaaattt gatttttaa aaaaatgtaa tagtcacata ttaaaaattc aaaggataca 26280
gaagatggag ggtttttgaa aacaagggca ttctttttt ttttctttct ttttttttt 26340
ttttttttg agatggagtc ttgttctgtc acccaggcta gagtgcagtg gcgccatctc 26400
ggcccagtac aacctccgcc tcctgggttc aagcgattct cccgcctcag cctcctgagt 26460
agctgggatt acaggcaccc accatcatgc ctggctaatt tttgtatctt tgtggagatg 26520
ggatttcacc atgttggcca ggatgaactc ctgacctcgg gtaatccacc cgcctcggcc 26580
tcccaaagtt ctgggattac aggcgtgagc caccacaccc ggccaacagc attctgatga 26640
tatagctggt gatagcagat gggaggacag tggtgtgccc agaaacctct ggacctggag 26700
cacaaaaggc tcagggtgca ttcccagccc agcggattct ctgcctgcat ctgctaagga 26760
tgaatgactc agctttgggg gaattagttg tgagactttg gacttcaggg gcggggcccc 26820
aagaggcaat ggtgataaag ttgaatttga gcagggaagg tgccccgtca gctgtcatcc 26880
ttttccccag gaacatcatt atgtaagact cctgccttgt ggaacaggct gtgagttgct 26940

gctcttccat tcctcacagc catgtttaca agggtaagga agggaaggag tggttcatca 27000
ttgcagagag gaaggtgcct tggccaagca gacctgctct gtgccaggca tgacactggg 27060
caaatgcaca gtatttagtt tatttatcct gaaatgcttt cagaactcaa tgcatccagt 27120
gtcttgttat ccctccagcc tatccgcaaa cctgctgaga tgcagtaggt ttggcgtaga 27180
atgcactgag ggtatctgtg gcaacagtgg gctaaagaac aaggcacatc aaggggtct 27240
tccgacgaac accccaaggg ttgcctccac cacgcagcat cctgctgtgg ctcgcctcaa 27300
tgtcccaggt gctctgtggg cacagtggcc aggtcagacc atgatggcca ctttctgact 27360
gtatggtctc atacagggaa ggcatgtcac ttttcttggc cttctctagg ttctcacctg 27420
taagtggggg taaaatgtcc cctccaaggg ttcctgtagg ggccgaagtg gctcaggcac 27480
ctggcacgcc tgttgcagcc cagctctgtg ctagcacctc ccaatgcctg ttgaatccac 27540
cattgccttc tgggacgtgt ctccatactt ccacgagaat acgtctaggg cacagcctgg 27600
ggttctgcat ttgagttctc acctccgtcc caggtgagcc caaaggtgct ggtctctagt 27660
ccacatttga gaggcaaggc tttaatatcc accacacaca tttattttgc agattggcaa 27720
aagcttagat taactagcca gcccaatgtc acttggctaa gtggtagagg ggaggtgctg 27780
tgtctgtcag actctagtct ggaattggag gtgggatacc taggttcaag tcctggatat 27840
gaaacttccc tgtcacatcc cttctctgag cccaacactt aatccgaaag tcatggtgac 27900
gtgggaggcc aggtaaagta ggagatgtct agctagattg gaatttcaaa taatgaggaa 27960
ttttacagca taaccgtgtc ctgtccaata tttgggacat atgctaaaac agatattcac 28020
tgtttctctg aaattccagt atagctgggc atcctgcttt ttatttgcta aatgtgcaac 28080
cctaggttgt gagacctctc tgacccacgc gtggcctcct ccaggaagtc ttcctgagcg 28140
ccagcctggg ctgggcatcc tcctctgtgt gtctgatgct gctctgacct catgaagctc 28200
taattgctgg ggcctgtccc caccttgatg tgggagctct tcggaggcaa ggaccatgcc 28260
aagttgcata tctgtgtgtt cctgagtcca gtggttcatt aaaagctttt ccctgagagt 28320
atccttaatg ccccggaggt aattctcttt tcacaacctt tctactgcct gaggctcttg 28380
aggactaatt ccagttaaaa gcagagggaa ggatgtggta ggaatagcac cgcatggagc 28440
tggactctgt gccccccgtg cagcaggcag gaccctccct ctgtgtcacc tccatgactc 28500
agggctcaga caggaagccc tcatcctcgt cctccacggg tctcatgttt gtcaaggcca 28560
ggggtatcag gcgtagtagg caggggccca tgcctgctcc catctgggag accctgcttc 28620
aggcccctct catggcccct tttgttcccc acagctcact gtacaacttg ttggttcaag 28680
gttagaaaat gcagttttgt gttgaggggg accatcacag aagacaaagg gtccaggatg 28740
aaggctaccc atcagcttgc agggctggga cgaatgtgag aagcaactga tgcttgtaca 28800
gtagagagta agcatgtagg gccagtcccc agaccttgcc tcccctcagc cttgacatgt 28860

gatctccatt tctggtggct acccctgcta gtggtctggc ttcaccatct tagccctgcc 28920
caggctaaga ccctcttcca tcagaacctg cagctgggat gctgggagca acagtcaggg 28980
cagagctgcc catcctccca ttcagggagc cctcaggaag tacttgggac cccccgaccc 29040
tttatagatt cagcctgcct catccccctcc atggaccaac acgcccttct cctcagcagt 29100
gggctggggg accaggctcc tgaactgctt gtggctgttc cagcagtggg gagatggagg 29160
gtcacacagt cctgagtcta tggctttgac agcaacgggt cctgactgca gctgtattcg 29220
tgaagcgaag tacctaatac aatcaccgaa atgtacaaat tggaccccta taggttcaag 29280
gattcttggt gtaggaggta tggcccccgc cccgggaacc aggacctcag cttttagaag 29340
caaaatgcat gaatgcagtg atggttaggc caagtgccaa gggagacagc caacccctgc 29400
tatcatggcc agaggaggga gagaccactc aggcctgggt ggtggtagaa atcctcactg 29460
ccaaggagat tgtatgcgca ggcgttggga aggcccggag aagcctgaga gacaggcttg 29520
gcttggttat agcagagctg ggtggaggga gcagattagt tggcttagca caggcttcct 29580
gcagggtggt ggttcttggc cagatttgcc ccagtggggc ttttgggaaa gtatagaaat 29640
attgttgttt ttttttttt gagacagagt cttgctccgt tgcccaggct ggagtgcagt 29700
ggtgtgttct tggctcacta caacctccat ccacctcccg ggttcaagtg attctcctgc 29760
ctcagcctgc caagtagttg ggactacagg catgtgccac cacacccagc taatctttgt 29820
atttttagga gatggggttt caccatgttg gccaggctgg tcatgaactc ctaacctcaa 29880
gtgatctgcc cgcctcagcc tcccaaagtg ctgggattac aggtgtgagc cactgtgcca 29940
gccaagaaac attttaggtt atcatatctg ggcaatggat gctactggct ttaggtggga 30000
agaggccggg gatactgtta aaatcatgca atgtacaagg cagcccccac aagagagttc 30060
tgtggttgaa aatgtccgta gtgttgaggt tgaggactct gctgtggggc aacagtagga 30120
gaaggggtgc taatagtcag gtggtggaca gcagggaatt acaggtacat cagttaggag 30180
tgtatacagc tgcaagtaag agaccaccag atggcaggaa cagtgggaac agaatggttt 30240
atctttttca tgtgtcaaga taagtggtgc taaagtcagg catggtggcc ctcacttata 30300
atcccagcaa ttcaggaggc tgaagagtga ggatcacttg aggccaggag ttcaagacca 30360
gcctgggcaa cacagtgaaa tgccatctct aaaaataaaa attaaaaaaa ttagccaggg 30420
gctgggtgca gtggctcacg cctgtaatcc cagcactttg ggaggctaag gcgggtggat 30480
cacctgaggt caggagttcg agaccagcct ggccaacatg gtgaaaccct gtctctacta 30540
aaaatataaa attagctggg catggtggca cacctgtaat tctacttact cgggaggctg 30600
aggcaaggga atcacttaga actggggagg cgtaagttgc agtgagctga gtcacgccat 30660
tgcactccag cctgggcgac agatcaagac cctgtcaagg aaaggaaagg agaggggagg 30720

ggaggggagg ggggaggggg gaggggggag gagaaggagg gaggggaag attagctagg 30780
catggtgatg agcacctgta gtccctccta gctactccga aggctacggt gggagaactg 30840
cttgagcctg ggaggtcaag gctgcagtta gtgatgatcg tgtcactgca ctccagcctg 30900
ggtgaaaaag tgagaccctg tctcaaaaag aaaaaagaca gataggaaga aagaaagaag 30960
aaaagaaaga aaaagagaga gagagagaga gggagggagg gagggaaagg aaaggaagga 31020
aggaaatgca tctgattttt gtgtattgat tttgtatcct ataattttgc caagttcatt 31080
tattagttct agtaattttt tttattaaaa aaaattttca agatagggtc tcaccctgtt 31140
gtctaggctg gagtgcagtg gcacggttat agctcactgc agtctccatt gccaggactc 31200
aaacagtcct cctgcctcag cctcctgaat agctgggact acaggcatgc cagcatgcct 31260
ggctaattat tttatttttt gtagagatgg ggtctcacgt tgttgcccag gctggtctta 31320
aactcctggg ctcaagtgat tgtcctgcct cagtctccca aagtgctggg attataggca 31380
tgctccacca cactcagaca agttataata cattttcagt ggcgtattta ctgttttaga 31440
atataaaatc tatctgcaaa tagagataag tttaattttt ttctaatttg gacactcttt 31500
tccttcctcc ctccccttcc cctttccctt ccccttccct tttccctccc tccctctctc 31560
cctctctttc tctctctctc tttctctctt ttctcttttc ttttcatttc atttttgcca 31620
aattgctctg gttagaactt ttaacactat gttgaataga agtggtgaca gtatctcatt 31680
cctagacact tttcgaaaga agacatacat gcaaccaaca aacatttgaa aaaaaactca 31740
atatcactga tcattaaaga aacagaaatc aaaaccacaa tgagatacca tctcacatca 31800
gccagaatgg ctattattaa aaagtaaaaa aaagaaaaaa taacatgctg gcaaggtcat 31860
agagaaaagg gaacacttgt acagtgttgg tgggagtgta aattaggtca accattgtgg 31920
aaagcagcgt ggcaattcct cagagaccta aaggcagaac taccattcga cccagcaatc 31980
ccattactgg gtatataccc aaaggaatgt aagttgttct gccataagga cacatgcaca 32040
cgtctgttca ttgcagcact attcacaata gtaaagacat ggaatcaacc taaatgccca 32100
tcaatgacag attggataaa gaaaatgtac atatatgtca tggaatacta ttcagacata 32160
aaaaaagaca tgtgattatg tcctttgcag gaacatggat ggagctggag gccattatcc 32220
ttagcaaact aacgcaggaa cagaaaacaa aataccacat attctcactt atattctctc 32280
actaaatgat gagaactcat aggcactaag aggggaacga cagatgctgg aacccagtgg 32340
agggtgggag gagggagagg agcaggaaaa ataactattg ggtactagac ttagtacctg 32400
ggcgagaaaa taatctgtac aacaaacccc cgtgacacaa gtttacctat ataacaaacc 32460
tgcatatgta caacttaatg taaaataaaa gttaaaaaac aaggccaggc atggtagctc 32520
atgcctgtaa tcccagcctt ttgggaggcc gaggtgggcg gatcacttga ggtcaggagt 32580
ttgagaccag cctgcccaac ttggtgaaac cctgtctcta ccaaaaaagg aaaaaaaaaa 32640

aaaaagccag gtgtggtggt ccatgcctgt aatcccagct actcaggagg ctgaggcagg 32700
agaattgctt gaacttggga ggcagagttt gcagtaggct gagatcgctc cactgcactc 32760
cagtttgggt gacaaagcga gactctgtct caaaaaaaca aaaaacaaag ttaaaaaaca 32820
aaacatcgga caccacacac cacatggcag gatccaggat ccaatcagat caagctctgg 32880
catcacccca cggcaggatc cagtcagata ttaccttcca gcatcacctc attgtgagat 32940
ccaattagat catgcctcat tattaccctg tgcttataaa acccaaccca acccctagct 33000
caggaaaaga gattgagcat tccctccttc cttgccagtt gactttaaat aaagcttttc 33060
ttatctcaaa atataaaaaa gaaagtatct cccctgggca tggtgggctc gtgccggtaa 33120
tcccagcact ctgagaggca gaagtgggca gatcaactga ggtcaggagt tcaagaccag 33180
cctggccaac atagcaaaac cctgtctcta ctaaaaatac aaaaattagc caggtgtggt 33240
gcctggctaa tttccacgcc cggctaattt ttgcattttt agtagagacg gggttttgcc 33300
atgttggtta ggctggcctt gaacttctga ccttgtgatc cacccacctc agcttcccaa 33360
agtgctagca ttacaggcat gagccaccac ccccagcccc cttctgcctg atcttagagg 33420
aaaaaccttc agtctttcat cattaaaaaa aaaaattatt tttcgagaca gagtcttcct 33480
ctgttttcca ggctggagtg cagtgatgta atcgtggttc acagcagcct caaactcctg 33540
ggctcaagtg atcctcctgt ctgagcctcc tgagtaacta ggactacagg catgcaccac 33600
tacaccaaga tttttttgg tagggtcttg ctttgacctt cctttgacct tgctttgacc 33660
cttgatttga ccttgctttg acagtgtctt gtaatgttgc ccaggcttct cttgaactcc 33720
tgggctccag tgattctacc acattggcct cccaaagcag tgggattatg agcatgaatc 33780
attgagcctg ccagccttct gtcactgagg atgatataaa ctgtggggtt ttttggttgt 33840
ttttgttttt gagacgaaat ctcactctgt cgcccaggct ggagtgcaat ggcacaatct 33900
cagctgactg caacctctgc ctcctgagtt caagtgattc tcctgtctca gcctcccgag 33960
tagatgggat tacaggcgtg tgcaaccacg cctggctaat tttttgtatt tttagtagag 34020
atggggtatc accatgttgg ccaggctggt ctcttaactc ctgacctcaa gtgatctacc 34080
cgcctcagcc tcccaaagtg ctgagattac aggcatgagc caccacacct ggacattttt 34140
tttcatacat ggcctttatc atgttgagag agttacctgt attccttgtt ttctgagtgg 34200
ttttattatg aaaggatgtc ggatattgtc agatgtcttt tctgcatcgg ttgagagaat 34260
catgtgattt tttcccttca tcctgttaat ctggtatagt tcattaattg atttccatat 34320
gttgaaccat ccttatattc caggaataaa gtctacctgg tcatgatgta tactcttttt 34380
ttgttttgtt tttttttgga gagggagtct tgctctgtgg cccaggctgg agtccagttg 34440
catgatctca gctcattgca acctctgcct cccaggtcca agtgattctt ctgcctcagc 34500

ctcctaagta gctgggacta caggcatgta ccaccacagc cggctagttt ttgtattttt 34560
agtagagacg aggtttcacc atgttggcca ggctggtctc gaactcctga cctcaagtga 34620
tctgcctgcc tcggcctccc aaagtactgg gattacaggc ttgagccact gcgcctggcc 34680
aatgtgtata atctttttaa tacgatgttc agcttggttc gctagtactt ttactcagta 34740
ttcatgtata ttttattcaa tatttatgag atctgtagtt ttcttgtagt gcctttggtt 34800
ttgatatcag tataccatag gatcaggata ctatgaacat gccctcatag aataagttag 34860
gaagtgttct ttcctcttca atttagggaa gaatttgagg aggattgata ttatttcttt 34920
ttctttttct ttttctttct ttttttttt tgagatggag ttttgctctt gttgcccagg 34980
ctggagtgca atagcgtgat cttggctcac agcaacctct gccaactggg ttcaagcgat 35040
tctcctgccc cagcttcctg agtagctggg attacagaca tgtgccacca tgcccagcta 35100
attttgtatt tttagtagag acggggtttc tccatgttgg tcaggctggt ctcaaattcc 35160
gacctcaggt gatccgcctg cctcagcctc ctaaagtgct gggattacag gcgttgagcc 35220
accatgccca gctgatatta attctgcttt aaatgtttgc tagaattcgc cagtgaagcc 35280
atctgatcct gggcttttct tttgggggag tttaaaaatt actgattcaa tttccttact 35340
agttatatgt ctatttagat tttctgcttc ttcatgaatc agttttggta tgcaatgtct 35400
agcaatttgt ccatttcttc tagattatcg tttgttatac agtcatttat agtattatat 35460
tgtatttttt atttctgtaa aattgtaaag ttcccacttt catttgtgat tttggtaata 35520
tgagtcttct ctttttctta gtcaccttac ctaaaggttt gtcaattttg ttgatttttt 35580
tcttttaaaa tttatttttc tgtattattt tatttgtatt atagattgct ccttcagaat 35640
atttttttca agaaatcaac tttttgtttc atagctgttc tctatagttt tctattccct 35700
atttcactta tctcagttgt agtctttatt atttttaaaa attctagctt tgagtttagt 35760
ttttcttttt ctagctcctt aaggtgtgca gttagaaaat ttcaaatctt tcttcttctc 35820
ttctcctccc cctcctcctt cttcctcctc gcccttcctc ctcctccccc tcctcctgtt 35880
ggggtgatca gacccaacac caggtcgtgg gggtgacaaa gtccggtgga gtcaaaggat 35940
tgagacaaag acagtttgag agataaaggt gggacaccaa ggggccatcg tgatcatgga 36000
ggctgcgaaa gccctgcgct ctgggagtcc acagtattta ttggtaatcc aacaaagaaa 36060
caggtggtga ggcatgttct cactcatagg tgggaattga acaatgagaa cacttggaca 36120
caggaagggg aacatcacac accggggccc gttgtggcgt gaggggaggg atagcattag 36180
gagatatact taatgtaaat gacgacttaa tgggtgcagc acaccaacat ggcacatgta 36240
tacatatgta acaaacctgc acgttgtgca catgtaccct agaacttaaa gtattaaaaa 36300
aaaaaaaaag aaacaggtgg tgagaatgtg gaggtcaaaa gggcaggcgc atgatctaca 36360
gctgtgacag tttagcattt atatggaaca tgttctgcta cttgagataa tgggaatagg 36420

agcctaggag ggctagaagc aaggagccag caagtctaga cacattccag aggacattat 36480
gcaagtcctg cctcagtttc cctcccaaca ctcagctttt tcccaacatc ctcctcctcc 36540
ttcttctttt tttctttctc ttcctcttcc tttctttcct ccttctctct tttttgtaga 36600
gatggggttt tgctatgtta atccaggttg gtcttgaact cctggcctaa tatgattctc 36660
ctgccatgga ctcccaaagt gttgagatta caggcatgag ccaccacacc tggccctttc 36720
tttaaaaaaa ttttttttt ttaattttta aaaatttttt tgagacaagg tcttgctctg 36780
ctgcccaagc tggagtacag tgctgtaatc tcagctcacc gtagcctcga cctcttgggc 36840
tcaagtgatt ctcatccctc agccttccaa gttactggga ctacaggcac gtgccaccat 36900
gcctggcgaa tttttcctat ctttcttgta gagacagggt ttagccatgt tgcccgagct 36960
ggtctccctc aatcctgccc ccttggcctc ccaaagcact gggactacag gcatgatcca 37020
ccgcgccagg gtgctttctt ctttttgat tgtgtttatg gctaaaaatt ttcctcttag 37080
cacagctttg ctgcatccca taagttttgg tatgttgtgt tttcattttc atttgtctca 37140
aggtattttt atatttcctt tgtgattttt gctttgatcc attggttgtt aagcatgtgt 37200
tgtttaattt ccaaatatca tgaattttca gggtttttt cctgtaattt atttactttt 37260
ttttttttg agccaggatg gagtgcagtg gtgtgatcat ggctcactga agcactgatc 37320
tcctgggttc aagtagttct ctcgcctcag ccttctgagt agctggtacc ataggtgtgt 37380
gccaccatgc ctggctaatt tttgttttga aacagggtct cactctgtgg cccaggctgg 37440
agtgcagtgg tgcgatcatg gctcactgca gccttgatct cctaggttca ggtgatcctc 37500
ccacctcagc ctcctgggaa gctgggacta caggtgcaca ccactacacc agctaatttt 37560
ttgtattttt agtagggatg ggatgtcacc atgttgccca ggaagttctg aactcttggg 37620
ctcaagcagt tcatttccct cagcctccca aagtattggg attacaggtg tgagccacca 37680
cacccaactt atttttattt ttagagatgg ggtttctcta tgtttcccag gctgatcttg 37740
aacccctggg cccaagagat cctcccaact tctcctccca aagtgctgtg attacaggtg 37800
tgagtcaccg tgctcagccc cttctattat cgagttctag tttcattcca ttgtgactgg 37860
aaaatatact ttctatgatt ttaattattt aaaatataac aaggcttgtt ttgtcgccta 37920
acgtactgtc tgtcctgagg aatattccat atgcacttga aagaaatgtg tatcctgctg 37980
ttatggagtg gaatgttgta tatatacaag tgtccaagtg ttttataaat gttcaagact 38040
tctatttcct tactggtctt gtggctagtt gttccatcaa ttattgaaaa tggagtattg 38100
aagtctccaa ctacttattg ttgcattgtc tatttctcct ttcaatgatg taatgtttgc 38160
tttacatatt ttaaggtcac attgtttggt gcatatatat tattacttat tcttgatgaa 38220
ttgacccttt tagtaatgta taatgtcatt tttgtctttt gtaacaattc tttatttaaa 38280

ttctattttg tggtcaggtg caatggctca tgcctgtaat cccagcactt tgggaagctg 38340
aggtgggcag atcacttgag gtcaggagtt caagaccagt ctgtccaaca tggcaaaacc 38400
ccgtctctac taaaaataca aaaaattagc tgggtgtggt gggacacgcc tgtaatccca 38460
gctgcttggg aggctgaggc acaagaatag cttgaacccg ggagacagag gttgcagtga 38520
gccaagattg tggcactgca ctccagcctg gacaacagtg agaccctgtc tccaaaaata 38580
aataaaataa aaattctatt ttgtcagata ttagtgtagc aactccagct ctcttttggt 38640
gactatttgc gtggaatatc tttttctatt cttttatttt caaactattt gtgtcctcag 38700
atctaaagtg agtgtcttag acatcatata gttggatcct atctctaaaa caatgtattc 38760
tgcattctcc aactttgact acagagttga atccatttaa atttgaagta attactgata 38820
aggatttatg ccattttacc ttttcttttc tgtatgtctc atagattttt gtctttcatt 38880
ttcttcatta ttgacttctg tatttattta tttatttgct ttttgtttta ttttattaat 38940
tttttgtaga gacagaatct cactatgttg cccaggctgg tcttgaactc ctggcctcaa 39000
atgatcctcc tgcctcagcc tcccaaagtg ctgggattat agacatgagt caccttgctt 39060
ggctgggttt taaaaattgt ttttgtagtg acacattttg attctctttt catttccttt 39120
tgcatatatt ctatgtatta ttcttcgttg ttaccctggg gattacaaat aaaatcctag 39180
agttataaaa atctaatttg aattgatacc aacttaacag catacaaaac tctactccta 39240
tacagctttg tccctgcttt aggttattgg tgtcaaaaat tccatcttta cacattgttt 39300
gctcaaaaat atagaattat gtttttttgg ctgggtgtgg tggttcacac atgcaatctc 39360
agtgctttgg aaggttgagg tgggaggatt gctcgaggcc aggagtttga gaccagcctg 39420
ggcaacataa caagatccca gctttacaaa aaagggaaaa agaaagagtg acttggcagg 39480
catggtggct tagacctgta atgccagtac tttgaaagtc tgaggtggga gaattgcttg 39540
cctccaggag tttgggacca gcctgggcaa cacagtgaga ccccacctct acaaaaaata 39600
caagattttg ccaagcgtgg tggcatgtgc ctgtaatggg attccagcta tttgggaggc 39660
tgaaatggga ggatcagttg agcccagagg tcgaggctgc agtgagctgt gattgcacta 39720
atgcactcca gtctgtctca aaacaaacaa aaaacacccc aaaaaaaccc caaagttaaa 39780
ataattctgg cttttatatt tacctatgta aatacctta ttgaggattt ttatttcttc 39840
aaacatcttt gagttactgt ctagcatcct ttaatttcaa cctgaaagag tccccttagc 39900
atttcttata aggcgggtct agtggtaatg aactctctca gctattatgt atctgaaaat 39960
gtcttaattt ctcacttatt tttgaaggat agttttgctg aaataggatt tttggttgat 40020
aattttgttt cagcgcttta aatatatcat gctcactgcc ttttgacctc caatgtttct 40080
tatgagtaat cagctataat cagctgataa tcttattgag gacaccttgt atgtgatgag 40140
tcacttttct cttgttttca atattctctc ttagtatttg cctttcaact gtttgattat 40200

aatatggctc aatataagtc tctttgtatt tatattcctt ggcgtttatt ggacttttca 40260
gatatttaat attcatgtct ttcatcaaat ttggaaagtt ttaggccatt atttcttcaa 40320
ataatctgtc tcattctccc tttcttctcc ttattgaact cccataacac ccacgttatt 40380
ttgtttcatg gtgtaccata agtagcagtc tctgttcact tttcctcact ctttttcatg 40440
tctgttcctc agacctgatg atttcaattg tcctaccttc aggttcacag attctttctt 40500
ctgctttctc aaatctgctc ttgagcccct ctagtgaagt tttttatttc agttattatc 40560
cttttcaggt ccagaatttc tgtgtggttc ttttttataa tttctcttta ttgatatcct 40620
cattttgttc atgcatagtt ttcctaattt actttagtcc tcatccattt ttgctcttag 40680
ctctttaaga tagctatttt aaagtttttt gtctaataag tgtaatgttg ggctgccttg 40740
gacacagttt ttgtcaactt ttttttttt ttcctttgaa taggccatct tttcccattt 40800
gtctgacttg tgattttgct gttgctgttg aaaactggac atttgactat tataatgtga 40860
taagtctgga aatcagattc tctctcttcc tcagcatttt tttttaattt ctgaagactg 40920
tagtaatgtt tgtttttata ctttcccaag ctatttttgc aaagactatt cattgttttt 40980
ttgtggtcac caaagtgtct gtttcttcag cttgtgttta gccagtgttt tgacagagat 41040
ttccttgaat gccaggagct aaaaaacaac accaacacac acacacacac acacacacac 41100
acacacacgt acacacacaa gcatacctct cctatctttt gcaaattggg gttgggactc 41160
ttttaacact tagctaggct tgttctgagc ctaggatcag cctgcgacaa aagtttcagg 41220
gctttttctga acatgtgttt tgccttgtac atgcatgcgg cattctcaat ttcctgtata 41280
catagccgtt ttatttttgt ttgagttgga tctcactctg tcgtctaggc tggtgtgcag 41340
tgacatgatc atggctcact gcagccttga actcctgggc tcaggtgatc ttcttgtttc 41400
tgtttctcga gtggctggga ctacaggaat gcaccaccat gcccagctaa gtttcccttc 41460
ccttcccttt tctcctctcc ccttcctttc ccttcctcct ttcttttctc ttttcttttc 41520
tctctctttc tttcctttte tttcttcttt ctttctttcc tttcttttte tttctttctt 41580
tctttctttc tttctctctc tctcctcccc cccttccttc cttctttcct tccttccttt 41640
ccttttttct tttgcttttt tctttcctct tctttcttct tttctttctt tctttctttc 41700
tttctttctt tctttctttc tttctttctt tctttctttc tttctttctt tctctttctc 41760
tctctctgtc tcctccctgc caccctcccc tccttccttc ctttcctttt ttcttttgct 41820
ttttttcttc ctcttctttc tttctttttc tttttctctt tctttttctc tctctctttc 41880
cttctctccc tccctccttc ccttcccccc ccctcccccc ctcctccccc cccctcccca 41940
tcctgtcctt gtgtgaacat agctcacagc agccttaacc ttgagggctc aagtgatctt 42000
cctgtgtctc ttccaagtag ctgggacagc aggtgcctaa cctccgtcta attatttatt 42060

ttttttctgct catcctctgt gggttggacc cactgccgaa ccagtcccaa tgagatgaac 42120
tgggtacctc agttggaaat gcagaaatca cccaccttct gcactggtct cactggaagc 42180
tgcagatggg aactgttcct attcggccat cttggcccct tccaattatt tatttttttg 42240
tagagacagg gtctcatcat gtttcccagg ctggtttcaa actcctggga tcagggcagg 42300
atcttcccac ctcaacctcc caaattgctg agattacagg tgtgagccac catgcccagc 42360
ctgcttttct attttgttgt agagacaggc tctcactacc ttggccaggc ttgtctcaaa 42420
ctcctggcct caagcagtcc tcttgccttg gtcttccaaa ctgctgtgat tacaggcatg 42480
agccactgca cctggcggct tcttcttctt ctttttttttt ttcttttgag tcaatgtcca 42540
gcctggagtg caatggtgcg gtatggctta ctgcagcctc aaacccctaa actcagatga 42600
tcctcccacc tcagcctccc aaatagctgg gactacaggt acatgccacc atgccagcta 42660
actttttta cattttattt tttgtagaga tgggggtctt gcaattattg cccaggctgg 42720
tctcaaactc ctggcctcaa gtgatcctcc caccttggcc tccaaaagca ttgggattac 42780
aggcatgagc cactgtgctt ggctcaaagc tgctttaaaa atttatgtac atatatatat 42840
tttaagacag agacttgctc tactgcactg gctgtagtgc agtggcacaa tcatggctca 42900
ctgcagtctc aaacttctgg gctaaagcaa tcctcccgct tcagcctccc aagtagctgg 42960
gactacagtt gcatgccacc acccccagct aatttttaaa ttttttgtag agacagggtc 43020
ttgctatgtt gtccagactg gtctcaaact cctgggctca agcaatctgc ctgcttcagc 43080
atccgcaagt gttggggtta cagatgtaag ccactgcgcc cacgagttgc tgctgaatat 43140
ccaaattgtc taagcttctc ctctgggttt aaaatggtct atggcatgtc tctacctata 43200
acctcttgcc ccaggcatct tttctgagca atgtcctgat tttaggtaag agatacagca 43260
tcttgcatca gttcttccag gatcccccag acaagaacag atgcacgtaa tagtttgcaa 43320
ataaggcctg ctctctttgg aggagggagc tgagaactgt actactgttg tctcaattcc 43380
aaaactgttg actgagtgca gtggctcacg cctgtaatcc caacactttg ggaggccaag 43440
gcaggaggat cacttgaggc caggagtttg agaccagccc agacaacata gtgagaccct 43500
atctctacaa acaatttaaa acactagctg ggtgtggtgg cacatacatg taattctagc 43560
ttctcaggag acggaggttg gaggattgct tgagcccagg agtttgaggc tgcagtaagc 43620
catgattgta ccaatacatt ccagcctggg ctacagaatg agaccctgct tcaaaaagaa 43680
aaaaaaaaaa aaaagaccaa gactgctgcc atgctgggga aggggtgggg caagactaag 43740
taaaaacacc acaaaacttt gctactgttt tgaagatggc ctttttttaaa ttgagtgttt 43800
gcctggttgc tgtaggcctt tgttttctag agtgacaaca aagttggttc tgacagtttg 43860
gcttgtttat tcagtgtttc agtttggaaa tgagagcttg gagcttccta ggccaccatt 43920
ttgctgatgt catttccaat ggcatttttt gcatctcgac tttttcctcg cgttcaatgc 43980

ttcaggacca caagatggtt gctacagctc tagaccttcc atctgtctag tgtggcgaaa 44040
agtggggaag gctagaatat catgccagct gcatacctcc cctttcatga gggaagaaaa 44100
agccttccca cggggatcac agggcccctg ctagctgcaa aggggtctgg gagaacaggg 44160
agagcctctc tcacctgagc agtggacaca atccttcacc aaagagtgca ggttctgatg 44220
gcaagaaaga caaaggggcc accggcaggc tcgttacccc aaagagcgag aagtagggga 44280
tgtgattact tacatctgta ccagttagag tgttgtacac atatccagcc aaggtacctg 44340
tggcccaggt caggtgactg gcttagcaat ttcacctacc ttcctctcag cccagatccc 44400
caaattcttt gaatgctgtt gggatgcaga acagcaagtc agcgagtgat ttttttaat 44460
ttaattttta tgagtacaca gtagattata tatttatggg gtacatcaga tattttgata 44520
cagatataca atgtgtcata atcacatcag gttgtaaatg gagtgaccgt cacctcaagc 44580
atttgtcact tctctgttac aaacatttta attacaccct tttagttatt ttaaaatgta 44640
ctgctgattg taattaccct gttatgctat caaataccag ctcttattca ttctatctaa 44700
ttatattttt gtacccacca accatctccg cttcccccta cctccccact actcttccca 44760
gcctctggta accatcgttc tacctactgt ctattgccat gcgtttgttt tcatttttag 44820
ctcctataaa tgagtgaaaa cacatgaagt ttgtctttct gtgcctggct tatttcagtg 44880
agtgatcctc atgtctccag ggcttgtctg tacatgactc acctggggca gcctctgcca 44940
ggtgtcaccc cggagccagc aacaaagggc tgctctgctg atggctgcct caccccggc 45000
tgctccctca gtgaactggc acagctctgg gcccctctcg ggaccttctc agagtagcca 45060
catttcagac ctgtcttatg attctaacat caaacttata atatcaatct tactaatacc 45120
aatagaaagt ggaaaatgag gtattatctg gcagtcatta aattagtaag ttctaatgac 45180
aaacataata cacgatgaag gtgagactgt gggaagatgg tgcctttgcg agttgcccat 45240
gtcagtggta agagtcacgg ccctcgggaa atcaccgagt cttcattacc caagactggc 45300
atcaaccctt caccaaattc caataactga gaatctgata attacccaat aaatcctaga 45360
ttagcctgag gaaagaaatg agctgtccac gtaagagtcg taaacattgg gccggggctg 45420
gtggctcatg cctgtaatcc cagcactttg ggaggccgag gcaggcggat cacgagatca 45480
ggagatcgag accatcctgg ctaacatggt caaaccctgt ctctactaaa aatacaaaaa 45540
tgaaccaggc atggtggcac atgcctgtag tcccagctac tcgggaggct gaggcaggag 45600
aatcatttga acccaagagg cagaagttgc agtgagccga gattgcgcca ctgcactcca 45660
gcctggcaac agagtgagac tctgtctcaa aaaaaaaaaa aaaaaaaaaa gaatggtaaa 45720
cattgtactc tgactcacaa atctcatcta ggggaacttg ttttaaggaa ataaattcaa 45780
agaaggagga aacattgtta ggtgcaaaga agtcaaccag aaacttattt atcaaaaatg 45840

aactattggg aaccggctcg acagtcagca ccagaagagg agaagatcca cgcgttctgt 45900
ggaccataac ctagtcacgg acgtgctgat cagagattga aggcaacagg gaggatttat 45960
gtgaaaagtc aagagaaaaa gcaggatgca tgtacatatc atatggttac agctcggcac 46020
gtgtgtccag aggcaccggc agctgggttg ggagatcggg tgtgaaattt tcactgtcat 46080
tccgagccgg attgtgccgc tgttatgctg cgtgtgtttc acaaatgacc ccaggagacc 46140
acatagctgg actctatctc tctgtggtgc tagactgggc acagctgggc tccaggggct 46200
tagcctagac agcccccatg ggaagaaaca tatgaaaggc agggtgggcc tttcatatct 46260
ttgttctgac acagctctgt gcatgccgac agtgtcttct tgtcgcaagt gcccacggcc 46320
ctgcctaagg ccctttgaca ctgaaggtgc ccgccacgtg ctggggcgaa atcttccagg 46380
aatgtcctct accagtgaca gatgaatgtg gtggaaagct gtctgtgtcc ttattccttg 46440
gaggggacct tcttgggcac gtccccacca gttcccggag gtccctgggg gcaggagcaa 46500
gctcttggat gcattctggt cagctttctt ccatcccctg gctcattccc cattcaccga 46560
ctgctgtcat ctggggtcat ctccccaata aactctttgc actgggatcc ttgtttcagg 46620
atctgtttct ggaggaacta gatgacaaca ccgggaacag aggacctaga gaggcagctt 46680
catgggtggt ggggtgtccg cctctgccgg ccagggactt gggagcagtg ctgggaaggt 46740
gctggatgga gctgtcactc acaggggcag gtccttggct gctgactgtc ttcctctcca 46800
ctatggctgt cttgagaact taggggtcag cctgaccctg ccttggcccc cttcctctca 46860
gcctctgtct tctcctgcat gaggctgggt ggctcccctg tgaatcaggc aggggtccac 46920
agaacactag agacaggtcc cttcctgcag ctgtctccag taggtggcca cgcaggagat 46980
gttcccaaca agctgccctt atctgcagct cagctttggt aatgggggcc cattaccaaa 47040
tgggggtaaa ggtcatggcc catcctggtg atagtgagaa cccaaggtag gccttgaaga 47100
ttcctatcag gagggagcag aaagtgtgta ccacacccct gggcccaggt ggagcagggc 47160
tgctgctcaa ggctcccagc catgctctgt cccttgctag gggtgaccgg tgggacaggc 47220
ctgggcaagg gacaagaggg agaaggtcgg ggggaagagg ggatgaagag caaagtgagc 47280
aaaggagagt cttccactat ctggggtctc tgtcaactgt caggccctag agtgagctgt 47340
tctttccctt tgcttcctgg aggaggggac ttttgtcact gcgtcactcc accctgcctg 47400
cccctccgtt atcaggctgt taatattaat taacaacagt tgctagggat gacagtgcag 47460
agggttcctc tgagcccatt gctggccctg gtcccaagag ggggtagggc agagctgggg 47520
tctgaggctg agccagggag ggtgcggagg ttcctcggcc atgctgagct cctgaggccg 47580
ggtcccagcc agtgcctggt cccatctgtg cctccaggcc ctggcaccaa ctccagcagt 47640
gttaggggct aatagcgtgg tctctcccct agctgactca gccctctggc ttcggtcgct 47700
ttgggaagtg agtggagacc ctagcacctg cgtgatgagg ctcatctaaa gcgggggcct 47760

gtggactggg gccaaacagt gggagtggtg gatcattaac cagcagggct cagcctcatt 47820
ggtccctaac ccagtcaggc cagggttgtc atcgaagggg aggaggctgc cttaatgtgt 47880
gttcagccct tggctgttcc tgaggcctgg cctggctccc cgctgacccc ttcccagacc 47940
tgggatggcg gaggccggcc tgaggggctg gctgctgtgg gccctgctcc tgcgcttggt 48000
gagtcccagg gcttggctcc acctcccctg cggcctccag ttagggaccc tggggccagc 48060
cgtgtaccag gcgagcgtta ctgggtgaca gcaagggagc ctcagggcct gcgggctggg 48120
caagtctctg gacacatgag ggatgccagg ccccacagag gaggggtgca ggtggagggt 48180
ttccaggtta caggcttgaa tgcacacagg ggtgaaagag gctgctggac tggggtgctc 48240
caagtccctc ctgtcactgg ccctactgtg gggtccaggc ctgcagttga gggaggtctg 48300
aggcaaggag gtgctgggat ggggttacct ggtgagcatc acctagggag gactgagcac 48360
tctggaggct gggagaagat ccagcgctgg cacctcttaa gttcctcgct tactttgtgt 48420
ctgggaggtg ggtgacagct tttggcctca agcaggtggt ggtagtggtg gtgggagtcg 48480
gggggcctcc tgaacagact ctccatgaga gaccctggcc tctggatgtg gtgtacagtg 48540
tggggactca ggctgacttt gacgtgggca gagcccggga ccttggagtc agctttgcct 48600
ccttacccat ctctggcctc tccagcatga ctttcctaag ctgcaggtct atcaggccac 48660
ccccaggaag aaaggccagt gttgtcactc caacactggc tggctggcac atgcctccag 48720
gaggcttcct actccccaca ctccccgctt ccctgccccc gctccatgtc cttcttaccc 48780
tcacaccctc cctggctgcc tgctgcctgg atggcaccca gctgtgtcag ggcccacgcg 48840
tgatgttgct gtgctctgca ggcccagagt gagccttaca caaccatcca ccagcctggc 48900
tactgcgcct tctatgacga atgtgggaag aacccagagc tgtctggaag cctcatgaca 48960
ctctccaacg tgtcctgcct gtccaacacg ccggcccgca agatcacagg tgatcacctg 49020
atcctattac agaagatctg cccccgcctc tacaccggcc ccaacaccca agcctgctgc 49080
tccgccaagc agctggtatc actggaagcg agtctgtcga tcaccaaggc cctcctcacc 49140
cgctgcccag cctgctctga caattttgtg aacctgcact gccacaacac gtgcagcccc 49200
aatcagagcc tcttcatcaa tgtgacccgc gtggcccagc taggggctgg acaactccca 49260
gctgtggtgg cctatgaggc cttctaccag catagctttg ccgagcagag ctatgactcc 49320
tgcagccgtg tgcgcgtccc tgcagctgcc acgctggctg tgggcaccat gtgtggcgtg 49380
tatggctctg ccctttgcaa tgcccagcgc tggctcaact tccagggaga cacaggcaat 49440
ggtctggccc cactggacat caccttccac ctcttggagc ctggccaggc cgtggggagt 49500
gggattcagc ctctgaatga gggggttgca cgttgcaatg agtcccaagg tgacgacgtg 49560
gcgacctgct cctgccaaga ctgtgctgca tcctgtcctg ccatagcccg cccccaggcc 49620

ctcgactcca ccttctacct gggccagatg ccgggcagtc tggtcctcat catcatcctc 49680
tgctctgtct tcgctgtggt caccatcctg cttgtgggat tccgtgtggc ccccgccagg 49740
gacaaaagca agatggtgga ccccaagaag ggcaccagcc tctctgacaa gctcagcttc 49800
tccacccaca ccctccttgg ccagttcttc cagggctggg gcacgtgggt ggcttcgtgg 49860
cctctgacca tcttggtgct atctgtcatc ccggtggtgg ccttggcagc gggcctggtc 49920
tttacagaac tcactacgga ccccgtggag ctgtggtcgg cccccaacag ccaagcccgg 49980
agtgagaaag ctttccatga ccagcatttc ggccccttct tccgaaccaa ccaggtgatc 50040
ctgacggctc ctaaccggtc cagctacagg tatgactctc tgctgctggg gcccaagaac 50100
ttcagcggaa tcctggacct ggacttgctg ctggagctgc tagagctgca ggagaggctg 50160
cggcacctcc aggtatggtc gcccgaagca cagcgcaaca tctccctgca ggacatctgc 50220
tacgcccccc tcaatccgga caataccagt ctctacgact gctgcatcaa cagcctcctg 50280
cagtatttcc agaacaaccg cacgctcctg ctgctcacag ccaaccagac actgatgggg 50340
cagacctccc aagtcgactg gaaggaccat tttctgtact gtgccaagtg agtccatggt 50400
ggggcccaag cgaggagtgg gctggggctg gggctgggct gccatggcct cctgggaacc 50460
tggccgggca tacagctggt cctgaaggac cagaggtagc tattcctacg gctctggcct 50520
ggggccgccc agatgattat ctctgccccc cgtccggccg ccatttcctt tggtcagagt 50580
tcctgctcat ggctgcaggt ttgtgcgtgg ccatcgctgg cccttcaacc ccgagtccac 50640
tctgtctttc tgcagatttc ttgacatgtg ggagctcccc gccacactct tgctttaagt 50700
ctgacagagg agcccgattg gcagagtaca tatttatatt tgctatgttt tgcttcttgt 50760
ttctgtgcca ggggccgtag ggccatcagt aacccatgag gtaccatggt atgcattgga 50820
aaaggtgccc tcaggccaga ggtcgtggct ggtctcaggc acctgggccg ggtgtcctgg 50880
ggtaggccac agccacacac acttctattg attggggttc ggtctttggt tctgtccact 50940
ctggtgtgct gccaacaaga tgccaacaac gctgctgggc caaggggcc aagagccaag 51000
ggcagcagca gggccttggc agtggaggct ccttgaggtt ggagtagagc agaggtcctc 51060
aagatgaacg tttagtactc catactccag agcaaatgag agttaaaagg ggcaaatagc 51120
atcttagtgt tattatgaaa acagttctga ccttacagac cctggaaagg gtctccagga 51180
cgcctaaggg ccccaggcca cactttgaga accactggat tggaagagag tgccgacact 51240
ttctgtcccc tgctacctgg ctctgcatcc ctcagctggg ccccaagttt gggctgcttc 51300
ccagagtgtc tgtgccagga acccaagggc tctctcttgg aaatagcagg aacgagagga 51360
gccattgttt gctctgggga ggcatcatgg tctgacctca gactcatgtc tgacggtagc 51420
tttatagtcc attatagggt attatcttta ttttgacttc ggatgctcac aacaactctc 51480
gggtggtcca attatctcca ttttacagac aggaaaactg aggttcagag gggtgtggta 51540

agctgctcaa ggtcacacag caaccagcac tcgcttgctg agatctgaga gaggggggta 51600
gagagctttg ctcaggtgtc ccactgcatc ttcgcaatga cgggctttgc agaaagggct 51660
aagctgaagg acctacagac ttgcctgagg gcaccagtct agtaaactgt gaaaacattg 51720
gctgctgggc tccagggttc caaatctaac ctcaatacct aaagggtttc gggggcccta 51780
ggcaggagaa ggaggctgag agggcaacgt ttgagacagc ccatgccaga ccccatggct 51840
caaatcccag ctcttccacc ctcacgggac ttcaggtgtg acgctcaatc cagagtcaga 51900
taatgtcaga gccaggaagg tcaggccagt gtgtggagac atgagaggct cagagggaca 51960
ggtcccggag cagcccctgc ctgccacaga gaaggcactc agggcagctc caactcactc 52020
cgtgggtggg ggcctgcagg agatcttgct ggatgggagc catttaggac ccactcggct 52080
gggtcctaaa tagctaaatg gcctaaatgc agatagctgg gctatctgca gccagtgtcc 52140
cccaccccac cagctcaccc tccatagtgc tgtgggtctg gggtgggagg ggaagggagg 52200
ggccataggg actgggcagg gccaggaaag gccctttccc tttgcggtca tctccctcta 52260
gtgccccgct caccttcaag gatggcacag ccctggccct gagctgcatg gctgactacg 52320
gggcccctgt cttccccttc cttgccattg gggggtacaa aggtaagcta agtgggccct 52380
gagaggaagc caaggaagat gcagtattgg ggcaggaacc atagacggga gggtgggagt 52440
ggtgctgggg attctcgcgg cctgggggta gcctggcttc tggaagctgt aggccaaccc 52500
tgtcctgttt cctctctctg ccatctcctt tatcttctag tagtgttact caggcactgt 52560
ggtttttctg cctgggccca aaggtctcgc ctttggctga gagaagtggg gtgtaggagg 52620
taaggccatg tatcagatga ggaaggagtg ggggagaagg agcaaggggt gatgggaggg 52680
gtgcagctag atagggggag ggaatatagg ggtgcagctg gagggggagg gaggcacggg 52740
tgcagcagga agggtctgag tatttcttat cccaggaaag gactattctg aggcagaggc 52800
cctgatcatg acgttctccc tcaacaatta ccctgccggg gacccccgtc tggcccaggc 52860
caagctgtgg gaggaggcct tcttagagga aatgcgagcc ttccagcgtc ggatggctgg 52920
catgttccag gtcacgttca tggctgaggt aggggctgca gggtccctgg ctctgggggt 52980
gcaacccagg tggtcttggg tcagttcctg tgtccccatc ctggccctgg cccttcctaa 53040
gtgaccctgg gcagtggctg cctgctcaga acggggtgat tgtgatggct gttcttatag 53100
cctcacctgc gattataggg ggccatcagg ccctatgaca caacacacaa ttagtgccca 53160
gtgaccgagc tattgagagc tggcctggct gaagcaggca cggtcagtgg gggctggtcg 53220
ggtgtgtgtc cacagcgctc tctggaagac gagatcaatc gcaccacagc tgaagacctg 53280
cccatctttg ccaccagcta cattgtcata ttcctgtaca tctctctggc cctgggcagc 53340
tattccagct ggagccgagt gatggtgaga agcgggaggg acacagctaa gtgggctagc 53400

ccaggacccc aggcatcttc agtaggcctt ctacaacttt cctaaccaca gcacctcaga 53460
acagcaaagt ggacacaccc aagtggctgc cccaaagggt aatacctctt gcaagtgttc 53520
tgtgctgaaa ggtcaagagc aattttcttt tcttttcctt tctttttctt ctcttttctt 53580
tgcttttctt ttctcttctc ttttccctcc taccctctct ttctctttct tttctttctc 53640
tctctgtttc tctttttctc tctttctttc ttttgagaca gggtcttgct ctgttgccca 53700
ggctggagtg cagtggcatg atcttagctc actgcaacct caaaactcct gggcacaagt 53760
gatcctcctg cttcagcctc ccaagtagtt gggactatag gcacttgcca ttgtgcccag 53820
ctatttttt ttttttttg agacagagct ttgctcttgt tgcccaggct gcagtgtaac 53880
ggcgcgatct cggctcactg caacctccgc ctcctgggtt caacaattgt cctgcctcag 53940
cctcccgagt agctggcatt acaggcatgt gtcaccacgc ctggctaatt ttgtgttttt 54000
agtagagatg gggtttctcc atgttggtca gactggtctt gaactcctgg cctcaggtga 54060
tccgcccacc caaagtgctg ggattacatg cgtgagctac cacgtccggc catttttttt 54120
gttttgtagt ttttgtagag atggggtctc gcttttgcc taggctggtc tcaaactcct 54180
gggctcaagt gattcttcct catcagcctc ccaaaatgtt gagattacag gtgtgagcca 54240
gcacacctgg cctaagagca gttttctgtc tgttacatgc cataccctca cttgcccaaa 54300
tgcaaagcta agacttaaaa tctcttgcaa tgcatgctca aggaagatgg agtaggctca 54360
cccatgcctt tgggtttcct ggacctcccc ttgggaggat ggctctgcag aggggcttta 54420
atgtgagatg tgagctcctc accactgggg gcagtatcgg gcacctgcag gcactgaggg 54480
tgcctgccgg ctactttgtc tggcctagct gaggctggtg ggcatactgg gtaggtgcta 54540
agtggctagg gggctgagcc tgtttgcatt gcaggtggac tccaaggcca cgctgggcct 54600
cggcggggtg gccgtggtcc tgggagcagt catggctgcc atgggcttct tctcctactt 54660
gggtatccgc tcctccctgg tcatcctgca agtggttcct ttcctggtgc tgtccgtggg 54720
ggctgataac atcttcatct ttgttctcga gtaccaggta agaagggagg agctctccac 54780
acccccaact gcccactctt ctcccaacct cacctcctgg cctgatggga ctctggcgtg 54840
aatttgctgg gtctccctgc agactctttc tgttcatcga cacgcatgtt tacaatatct 54900
gtagaaacta gagtgtgttg acataaatga cttcatcctg cctctaccat ctggaattag 54960
ctttctgtta acccccttgca atgtctagta aaacctctcc atgttagtac attacagcct 55020
cctcctgtct ttatgctgct aggtagcatt ccatggtaag gataaatcag agtcgatttc 55080
acctctccct gttggtgaac aattagggtt ccaacagtgc ttggaacagg gatgctatag 55140
acatctcaaa tgcaccaacc atttctccca gccagaccct ggaagaagaa tattggccat 55200
ggagagtatg agagtctctg atgattcagg aaggtcagag cagctcctca ggcctggctg 55260
cagctctggg cacttgccaa ctccctgctg gcctttgagg ggcggtgccc ttggagggcc 55320

ctggctctta tccctgctgt tcccacacag aggctgcccc ggaggcctgg ggagccacga 55380
gaggtccaca ttgggcgagc cctaggcagg gtggctccca gcatgctgtt gtgcagcctc 55440
tctgaggcca tctgcttctt cctaggtgag cctgggtgag acctccccac tcggcattag 55500
gcttgctggg ttagtgccgg ggcctaggag ttcccagagg gcagtgggta tagtgcagat 55560
tcccttcccc ctgcaccctg tcaatgtcgg ctaccactct gcccttgaag ccagggtgcc 55620
ctgacagccc tctgctccct cacaggggcc ctgaccccca tgccagctgt gcggaccttt 55680
gccctgacct ctggccttgc agtgatcctt gacttcctcc tgcagatgtc agcctttgtg 55740
gccctgctct ccctggacag caagaggcag gaggtagggg cagctgggcc agtactgagg 55800
gacctgcccc tgggttccca ccatggcagg gagatggggt ggctttacca ccacagagat 55860
ggcccagaga atggggtggg ggacaggggc attgtgccag gagagtaata tttaggccat 55920
gtattctcca atttcctaca gaaaaataaa tttgttttga caatttttta aatataatca 55980
aacctcctaa agtgcatgat gttgagaaat aaaatacagt tgacccttga acaatgtgga 56040
gattagggca ccgactgtct aagcagttga aaatctgcat gtaacttttt ttttttttga 56100
gacggagttt cactctgtca cccaggctgg agtgcaatgg cgtgatatca gctcaccaca 56160
acctctgcct cccgggttca agcgattctc ctgcctcagc ctcccaatta ctgggattac 56220
aggccccctc ctcctgcacg cctggctaat ttttgtgttt ttaatagaga tggggtttca 56280
ccatgttggt caggttggtc tcgaactcct gacctcaggt gatctgccca ccttggcctc 56340
ccaaagtgct ggcgtgagcc accatgcctg gtctgcatgt aacatttgac ccttctaaac 56400
ttaattccta ctagcctact attgactgga agccttaatg ataacataaa tagtcgataa 56460
cacatctttt gaatgttata tgtattataa actgtattct tacaataaag gaagcaagaa 56520
aaaagaaaat gttagtaaga aaatcataag gaagagaaaa tctatttact attcacgaag 56580
tgaaagtgga tcatcatgag ggtcttcatc ctcgtcgtct tcaggttgag taggctgagg 56640
aagaggagga agaggagggc ttgatcttgc tgtttcaggg gcggcagagg tggaagagaa 56700
tccagggata agtgagccca ggcagttcaa actcgtgttg ttcaagggtc agctgtataa 56760
atgagaggtc gacaggagtt gatctgttgg ttcccatgat ggtgtaaaat ttaaagatat 56820
tttatcaaga ttaaaataaa agcaaagaaa acagcacact ggtatgtctc catgagggca 56880
ctggcacggg ccacccacag aaggtgacac tccctggggg caagaaggtg gtccctgggg 56940
ccttgtctgc tctgggacta ccttgagggg gtgcctccca ctccaggcct cccggttgga 57000
cgtctgctgc tgtgtcaagc cccaggagct gcccccgcct ggccagggag aggggctcct 57060
gcttggcttc ttccaaaagg cttatgcccc cttcctgctg cactggatca ctcgaggtgt 57120
tgtggtgagt gggcctcgaa ccacacgaga gcaggggcac taggtgggga cctcgcctca 57180

gggagagcag ggttggaggt ggggaggttg cctaggccca aatgctgata cttggggctg 57240
gcacgcaagt ctgctcaact ccagaatgtt gcccatgaca ccctgactga cttaaatttg 57300
tggggagatg ggggacggct gttgggcagg gtggtctcat gcagcaggtc ccttctcagc 57360
tgctgctgtt tctcgccctg ttcggagtga gcctctactc catgtgccac atcagcgtgg 57420
gactggacca ggagctggcc ctgcccaagg tgagcccagg ccctttctcaa cccttaggcc 57480
cctgggattt ggggaggggc agtagcaacc agcagggatg ggttgggggg tcctccggcc 57540
aggggcttgg ccagaggtgc agaattgttc attactctgg aggcacctcc agcagtcctg 57600
gggagtgaag ccacattcgt gtatgaacag cacaacagcc aggtgccagc cccaggccac 57660
agtaagagag atggcccagg catcggaggg ctgtccatgt gagatggcag gccacaaaga 57720
atgactgcca ctttgctgag tgcctgccca gtgtccagcc ctgcgaattc tctgggcctg 57780
aagcccgggg agggcagggg ttcaggggaa ggaaagcccc gtggttggag gggacctcca 57840
aggtcacata ggatttgcag aggaaagtga tgacagactc gccagtggga ggctagggtg 57900
agcccaggtg tgtttcctgg gcgtggcagc gactgtgggg gtgggatgag ctggaggcca 57960
agggcatggt cggggagagt gctgattgcc cagcctggac cagtaagtgt gcgggccaac 58020
aggcacaatg catcagccaa ggctggggac ccggctcctc tggatatgca tcagcggtgg 58080
ccatgggctg gtggccaaga ggaagcagcc acagacaaca aagtctgaga cacatggtca 58140
gactgcatga gcaagctcta gggagaggga aggcatcgag gggactcgat gtctaggtcc 58200
catctgggga actgtgatgg aggtttgggc aagggtctgg gtactggcag gagccccagt 58260
ggaagcagcc aggcctgagc ccacaacagg gctgagtggg gtgcggctgg ggtaggtgtg 58320
ttaggcagta ctggcctggg gtcctggaag ccaggtgagg gaggacaaga gcagatggct 58380
caggactgta ctttgggtga ctttatggag ggagagcagg tgaggagtca cagaatgaac 58440
ctgccacctg cagaagccct gggggctatg tcacagggct gaggtgaaga gggtctctag 58500
tgccccaaga gcaagaagga aggatgtgat gggctgccag accctgctga ggttttatgt 58560
tgatgtcttt tgtttatttt tctgttgggg acatttgttt cttactgctt ttaaaaattt 58620
tatcattttt tttccgtttt ttattgtggt aaaatacaca taatagaaaa ttaccattat 58680
aaccattttt aagtgtacag ttcagtgata ttaagtacac tcatactatt caactatcac 58740
caccatccat ctccaaaact ctttcctttt tgcaaaattg aaactttacc caacaaacag 58800
tgactcccca ttctcccctc ccctcagccc ctgacacaac cacctttttat ttatttattt 58860
attttgaaac agagtttcac tcttgttgcc caggctggag tgcaatggtg tgatctcggc 58920
tcaccgcaat ctccgcctcc cgggttcaag tgattctcct gcctcagcct cccaagtaac 58980
tgggattaca ggtggccgcc accacgccca gctaattttt gtatttttag tagagacagg 59040
gtttcaccat gttggcctgg ctggtcttga acttctgacc tcaggtgatc caccagccct 59100

ggcctcccaa agtgctggga ttacaggtgt gagccaccgc acctggcctc tacttttttct 59160
ttttttttt gagatggagt cttgctctgt cacccaggct ggagtgcaat ggtgcagact 59220
cggcttactg cagcctccac ctcccaggtt caagcgattc tcctgactca acctcctgag 59280
tagctgggac tacagccgtg tgccaccact cccagctaat ttttgtactt ttagtagaga 59340
cagggtttcg ccatgttggc caggctggtc tcgaactctg gaccttgtga tctgcctgct 59400
ttgccctccc aaagtgctgg gattacaggc atgaaccact gtgcccggcc catttacttt 59460
ctgttctatg agtttgacca ctctaggcac ctcaggtaag tgaactcata caatatttat 59520
ttttttggct gggagtggtg gctcactcct gtaatcccag cactttggga ggctgaggca 59580
ggcagatcac ctgaggtcag gagtttgaga ccagcttgac caacatggag aaaccccatc 59640
tctactaaaa atacaaagtt aactgggcat ggtggcacat gcctgtaatc ccagctactc 59700
aggaggctga ggcaggagaa tcacttgaac ctgggaggca gaggttgtgg taaactgaga 59760
tcacgccatt gcactccagc ctgggtaaca gagtgaggat tcgtctcaaa aaaaaaaaaa 59820
aagtatattt tgtctgatct tagtatagct acccctattc tcttttggtt actatttaca 59880
tggaatatct tttttctgtt cttccacttt caatctattt gtgtttttgg acctaaggtg 59940
agtctcttgg agacagcata tagttagatc acgttttgct gttttttagc agatgggggc 60000
tgcctagggc acagtatgct gactctcaca atctcgatcg tgtgtgtgtg tgtgtgtgtg 60060
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtttaattc attctaccac tctttttttt 60120
ctttttttt tttgagatgg agcctcagtc tgtcacccag gctggagtgc agtggagcga 60180
tctcagctca ctgcaactta cacctcccgg gttcaagcaa ttctcctgcc tcagcctcct 60240
gagtagctgg gattataggt gcatgccacc atgcctggct aatttttttg tatttttttgt 60300
agagacaggg tttcaccatg ttggccaggc tggcctcaaa ctcctgactt tgagtaatcc 60360
acccacctcg gcttcccaaa gtgctgggat tacaggcgtg agccaccatg cctggtccta 60420
ctactctctt ttgattggag agtttaatcc atttacattt acagtaatta ttgataagga 60480
gggatttact tctgtcattt tgctatttgt tttctatatg ccttgtagat tttttgtttc 60540
tcatttcctg cattactgac ttattttgtg cttagttgat tgctactagt gaaattttac 60600
attttccttc tcattttctt ttgtgcatag tctacagcta attttatttg tgattaccat 60660
ggggattatc ttaaatgtgc tgaagttata acactctaaa tttatgccaa ctttgtttcc 60720
atagcataca aaaactctgc cctataacaa ctccatctta cctccctttc agttattgat 60780
gtcacaaaat tatatcttga gctagccatg gtggcttatg cctgtaatcc caatgctttg 60840
agaggtggag gcaagaggat tgcttgaggc caggaatttg aggccagcct agccaacaca 60900
gtgagatccc atctctagaa aaaatttaaa atttagctgg gcaagatggc acgtgcctgt 60960

agtcccagct atgtgggagg cttgcttgag tccaggaatt caagtatgca gtcagctatg 61020
atcatgccac tgtactccag cctgagcaac agagagacac cttgtctcaa aaaatttttat 61080
ttttcagctg ggtgtagtgg ctcatggctg caatcccagc actttgtgag gtggttggat 61140
cacttgaggc caggaggtca agattagcct ggccaacatg acaaaacccc atctctacta 61200
aaaatacaaa aattagccag gcatggtggc acacaactgt aaacctagct acttgggagg 61260
ctgagacatg agaattgctt gaatctagga ggtagaggtt gcagtgagct gggatcgtac 61320
cactgtactc cagcttgggc gacagagcga gactatgtct caaaaacttt tgtattttta 61380
tgcattatgt atccaaaatc ataggctaat gatttttttt gcatgagtct cttaaatcat 61440
gtacaaaaag gtggagttat aaatcataac atttataact gcccatttat ttacctttgc 61500
cagggattta tttatttatt taaagaggca gagtcttgct ctgttgccca ggctgagatg 61560
cagtggtgtg atcatagctc actataacct caaactcctg gcctcaaaag atcctctcac 61620
ctcagccacc tgaagtactg ggattacagg tgtaagccac tatgcctagc caagggattt 61680
ttatttcttc atacatcttt gagttactgc tgatgtcttt ttttttttt ttttttttt 61740
ttgagaagtt gttttgctct tgttgcccac ccaggctgga gtgcagtggc atgatctcag 61800
ctcaccgcaa cctctgcctc ccgagttcaa gcgattctcc tgcctcagcc ccccgagtac 61860
tgggattaca ggcatgtgcc accacgccag gctaattttg tatttttagt agagatgggg 61920
tttctccatg ttggtcaggc tggtcttgaa ctcccaacct caggtgatct gcccgccttg 61980
gcctcccaaa gtgctgggat tacaggcatg agccaccatg cctggcctgt ctaatgtctt 62040
tttatttcaa cctacaggat tccttttagc atttctttca gggaaggtct agtgataacg 62100
aattccttca gcttttgttt atctgagaat gtcttaattt caccctcatt ttaatttttt 62160
aaaatttttt atttattttg agatggagtt tcactcttgt cgcccaggct ggagtacaat 62220
ggtgtgatct cagctcactg caacctctgc ctcctgggtt caagtgattc tcctgcctca 62280
gcctcctgag tagctgagat tacaggtgca tgccaccatg ccaggctaat ttttgtattt 62340
ttaatagaga cggggtttta ccatgttggc caggctggtc atgaactcct gacctcaggt 62400
gatccaccca ccttggcctc ccaaagtgct aggattacag gtgtgagcca ctgtgcccgg 62460
cccatttta ttttttaatt aaaacaattt ttttgagatg ggggtctcac tgtgttactc 62520
aggctggtct cgaacttttg ggctcaggtg atcctcgtgt ctcagcctcc caaagtattg 62580
ggattatagg acgaatcacc tcatctggca tctccctcat ttttatttttt aatttttagt 62640
ttttttttt ttttttgag atggagtctc actgtcaccc agattggagt gcggtggtgt 62700
gatctcggct cactgcaacc tccacctccc aggttcaaga gattctacta cctcagcctc 62760
caaagtagct gggattacag gtgcatgcct ccacgcctgg ctaatttttg tatttttagc 62820
agagatgggg tctcaccatg ttagtcaagc tggtctcaaa ctcctggcct caaataatct 62880

gtctgcctcg gcctcccaaa gtgctgggat tacaggcatg agccaccatg cctggccttc 62940
tccctcattt ttaagtgaca gttttgctgg aattaggatt cttcattgac aattgttttt 63000
tcttcagcac ttgttttttg ttgttgttgt ttgtttttga gacagagtct cactctgtca 63060
tccaggctgg agtgcagtgg catgatctca gctcactgca acctctgctt ctcaggttca 63120
agtgattctc ctgcttcatc ctcctgagta gctgggatta caggtttgtg ccaccatgcc 63180
tggctaattt ttgtattttc agtagagatg gggttttgcc atgttggcca ggctggtctc 63240
aaactcctga cctcaggtga tccacctgcc tcagcctcct gaagtgctgg gattacaggc 63300
atgagccatc atgcccagca ttcttcagca ctttcaatct acaaacccac tgccatctgg 63360
gcttcaaggt ttctgatgag aaatatgctg ataatctttt tgaggatctt ttgtatatgc 63420
caagtcactt ctttttttc aatattttct attttttaaa aaacttattt tattttactt 63480
tttattttta ttttttagag gcagggtctt gctatgttgc ctagaatgga cttgaaaccc 63540
tgggctcaag caatcctccc acctcagcct cttgagtagc tgggactaca ggtatatgcc 63600
accatgcctg gcttgtcttt ggtttttgac agctaaatta taatatccag ctgggtgcag 63660
tggcttatgc ctgtaatccc agcactttgg gaggccaagg tgggtggatc acaaggtcaa 63720
gagatcaaga ccatcctggc caacatggtg aaaccccatc tctactaaaa atacaaaaat 63780
tagctgggca tggttgtgcg cacttgtagt ccaagatact tgggaggctg aggcaggaga 63840
atcacttgaa cccaggaggc agaggttgca gtgagccgag attgtgccac tgcactccag 63900
cctggcaaca gagcaagact ccatctcaaa aaaaaaaaat tataatatcc gttggtgtgg 63960
gtttctttag tttatcctat tggagtttat tgagtttctt gaatgtttat attcatgtct 64020
ttcatcaaat ttggggagtt ctggccataa ttttttcaaa taatctcact tccccttttct 64080
cttttcttct ggaattctta caattcatat tttggtctat ttgatgatga tgatgtctga 64140
caggtccctt aggctctgct ctgttcactt tcgttatttt ttttcctttc tcttcttcag 64200
actcagtaat ttcaatggtc ttatcttcag tttgctaatt ctttcttctg actgcttttg 64260
aatccctcta gtgaattttt catttaagtt actgtacttt ttagctccag agtttttttg 64320
ctctttttta tgtttcctcc tcattgatat ttccattttg ttcataaatt tttccttgac 64380
tttgttttct tttagctctt tgagcaactt taaggcaatt gttttattca tttattttat 64440
tatttattta tttatttttt gagacagagt cttgctctat cacccaggct ggagtgcaat 64500
ggtgtgatct cgggtcactg caacctctgc ctcctggggt taagcctcag cctcccaagt 64560
agctgggatt acaggtgcct gccaccatgc ttggctaatt tttgtatttt tagtagagac 64620
agggtttcac catattagcc aggctggtct cgaactcctg gcctcatgtg atctgcctgc 64680
cttggccttc tgatgttgtg ggattacagg catcagccac tgtgcctggc tgagacaatt 64740

gttttgaagt ctttgtctag taagtctgct gtctggtctt acccaggaac agtttctgtt 64800
ggttaatatt ttccctttga atgggccatg tttttctttt tcttggtgtg tttttggttg 64860
aaaaatggac atttgattct tataatgtgg tagctctgga gatcagattc tccttctttc 64920
ccagggcttg ctttatttta tttattgctg ttggtgtttc tgtgctgggg atcagccaaa 64980
ggcacagagt taatgtcttc tcaggtattt ttgagactgc atttttctct gagcatttat 65040
gcagtgtggt gactgtctaa atatccctat atttatggtt gcttttgaat gtccttgtcc 65100
ttatatgtat ggttcccaaa aggagaaaaa gggaaaaatg aaggtgtcgg ggataggtgc 65160
ttactcttta aatctcctgg aagtcacttt agtaagatgt ggaggtggtt gcaacaacgg 65220
tggtgggagt tgcattagtg gctgcctgcc tgtgtatctg taccaccaat atcagaagta 65280
atgatcaatt atcagaactc agatccttga tatttgaact tatttattta tttattagag 65340
acagggtctg gctctcttgc tgaggctgaa gtgcagtggt gcaatcatag gtcactgcag 65400
cagcaaactt ccaggctcaa atgattctcc tatttcagcc tcctgagtag ctaggactac 65460
aggcatgtgc caccacaccc agctaacttt tgtatttttt tttgtagaga cagggtgtcg 65520
ctatgtgccc agatcggtct cccactcttg ggctcaagtg accctcctgc ttgccctccc 65580
aaagttctga aattacaagt gtaagccatc atgcccagct gatatttggt ggatggtgtc 65640
cttgcctacc tggctcctgc aagctgtgta caagctgctt ctggaaagca tacacagctg 65700
catgccttga ggctgggagt ggcaaatggg tagctgctac tgtactaaag ctgagattgc 65760
ctgaaattaa ccacaattta ctgtccaagc cttatcctgg aagcttccag ccctcaatag 65820
actccagagt tccaaaatcg ttacactagg gccggtgtgg tggctcatgc ctgtaatccc 65880
agccatttgg gaggccgaga cgggtggatc acttgaggtc aggagtttga gacaagcctg 65940
gccaacatgg tgaaacccca tctcttttaa aaatacaaaa atcagctggg agcggtggca 66000
catgcctgta atcctagcta ctcaggaggc tgaggcacaa gaatcgcttg aacccaggag 66060
gcggaggttg cagtgagcag agatcgcgcc actgcactcc agcccagaag actccatcca 66120
tctcaaaaca aaacaaaaca aaacaaaaac aaaatagtta taccagacaa attgttgtct 66180
agctggggag agggattcct gacacttcct actgtgccat tttccctaat gtcactctga 66240
gcctttatgt tatagaaggg agcagaccat gaggatgcct ggtgcatggc tttgagggtg 66300
tgcacactga catttatatg tgcacacaaa tatgggccgt tgtcacaggc cagcttgtta 66360
gacggtggct gtgccatatt gggggtgata ggaaggggta caattatgtg tctgtgcatg 66420
tttgtgtgtg tcagtgtgtg ttcatgtgag gtgataggtg ttgctctgtg tttgtacctg 66480
cataagtgta cttctgtttg cacctgtgat tatacctatt ctgtgaacct tggagtatgt 66540
tcatctgggg gtacacctaa aactgtgttc cggtgtaact gtacagtgca catacatctt 66600
gagggtaccc ctgagtgtgt gtgtctgtgc atgtccttct ctatatgtac cttgtgtgtg 66660

acctctgagc atgtacatct ctgtgtatat tttgtgtact tgtgtgcatg tacctctgtg 66720
tacctctaag catgtatcta cgtgtatatc tctgagtgtc ccactgagca catccctttg 66780
agtgtgtaac tgcatgtgtg tctctgaaca tgttcctctg tgtgttcctc tgatcatgga 66840
cctctgaaca tgtgcctttt agcatgtacc tctgtgtgta ccttcgagag tgtgagctgg 66900
attgagccct ttaggggtgt gcatagcgaa ccaaagctca ctgaccctcc tccactccta 66960
ggactcgtac ctgcttgact atttcctctt tctgaaccgc tacttcgagg tgggggcccc 67020
ggtgtacttt gttaccacct tgggctacaa cttctccagc gaggctggga tgaatgccat 67080
ctgctccagt gcaggctgca acaacttctc cttcacccag aagatccagt atgccacaga 67140
gttccctgag cagtgagttc ctggcccgcc ccaaacccca gcctactccc tgtttgagtc 67200
cctccagtcc tctccagtcc cctcttcctg atgttctatc cctgtcctgc tgccctgctg 67260
ccttgctgcc gtatgcctgg ggagggctgc gtgggggttg ggccacgaga aggacccacc 67320
accctgccca gctggccttt tcacccttcc tcccacctgc cccttaggtc ttacctggcc 67380
atccctgcct cctcctgggt ggatgacttc attgactggc tgaccccgtc ctcctgctgc 67440
cgcctttata tatctggccc caataaggac aagttctgcc cctcgaccgt cagtgagtgt 67500
ggggccatgg ggactcactg tccaccacag ctcgggcaaa ctgaggcaac agaaaggaga 67560
ggactggaga ggctccctca acctctccca cgcatcctgc agggtctgtc gggggcatgg 67620
gtgcagatgt ggcctgaggg acaggcactc tgtgagaagc acctgtgtgg gtgaccgtgc 67680
tggcccgtgg gcatcacaca tgtatactgc tgtgtactgt gcccccattt tcagagcaca 67740
tggtgctccc gggtggcagg gcagtgggga gtcaggaggg gagagctgct gaggttagca 67800
catggccctg ccgcccaaag cagtggcatt tgtaggtgga gaggcctttg tggggcctgt 67860
ttttctgccc caaacttcct ttccccttct gcctgtaggt gcccacagtt tctatagcca 67920
agaggagaac ttctcccaca aatgacaaat gcaaatcccc ctagaagcga ctggttgagg 67980
ctggagtgcc caggaccttt gatgggattg ttggggaagg aggggcacaa agcaggagct 68040
gctggccctg gggtgtcact gcccagaccc ctgctttctc tgcagactct ctgaactgcc 68100
taaagaactg catgagcatc acgatgggct ctgtgaggcc ctcggtggag cagttccata 68160
agtatcttcc ctggttcctg aacgaccggc ccaacatcaa atgtcccaaa gggtaagctt 68220
gggagggcct tctgctgggg aggacagaca tgtgggacac aggatggggt tgaatataga 68280
gaggcaggag gaggctatca ggggcctctc tggggtggct gtgggctggg cagatgaaag 68340
aagcttcgtc cctggctaag cctttgccct gaccttcttg cagcggcctg gcagcataca 68400
gcacctctgt gaacttgact tcagatggcc aggttttagg taagcatggc cttgcctgga 68460
ggggaggaca taaatcggtt gctctggagg gcccccgaaa accccaggga acagcctgtc 68520

acatgttgtc tccctccttt gtcaggaggt tctcactgcg ctggccctgt cagcaggggt 68580
cttgtttccc agctccacat ctcagacttc acccccttctc tcactcccaa gtccatggtc 68640
agtgctaagt ttgtggaatt gattcagcag ttgataccat acttgggagt tctccacacc 68700
ctggctaagc acctttctta ccagcacaaa ttacacccaa agggcagctg gttaaatgaa 68760
ttaggatgct tggcacagca caatcctagc agtcatttaa agtaacaaga ggctgggcgc 68820
ctgtaatctt agcactctag gaggccaagg cgagaggatc tcttgaatcc aggagtttga 68880
gaccagccca ggcaacagta gggagaccct ctttttttt tttcgagacg gagtctcgct 68940
ttgttgccca ggctggagtg cagtggtgca atctcggctc actgcaacct ccactttccg 69000
ggttcaagcg attctcctgc ctcagcctcc tgagtagctg ggactatagg agcataccat 69060
catgtctggg taatttttgt attttcagca gagatggaat tgcaccacgt tggccaagct 69120
ggtctcaaac tcctgacctc aggtgatatg cctgccttgg cctcccaaag tgctagtatt 69180
acaggcatga gccactgtgc ccggcctcct ctacaaagta aaatttaaaa aattgcccgg 69240
gtgtggtggc gtgtgcctgt agttccagct attcagaagg ctgggcggga agaatgcctg 69300
agtctgggag gttgaggctg tagtgaactg tgatcgcaac actgcactcc agcctgggca 69360
acaaagtgag accctctctc aaaaaaaaaa agaaagaaaa aagtaacaag agagatgcag 69420
ttggactgac aggaaaagga cccacaacat gctgtcagct tatacagcag atggcagaac 69480
aagacagcca tctgtgtaaa ggagctggcc atagctccgt gcagacatgc tcggtgtagg 69540
ggccctaagg gagctcgtgc tggagatgga catgggggtc gtcggtgggt gggggagttt 69600
ttgaaggatg atctcacttt gtactgaaat aattcatagt ttgaactgct ggctgaaagc 69660
tgcctcaagt tcgctcaccc cacccttcca gctatgaagt tcccatgttt ccagaagggc 69720
aatgcaccct gcccagccct ggtagctgag cacaacaggc tctgtgaggc cagtgtggtg 69780
gggctggtgt ggacagatgg gagtggatgt gtcagtcagg gaatgaggag cagggcctgg 69840
aaggagcaca cagtagagcc aagcccccat aaccgggggc aagtctgcac catctctgac 69900
ctttgtcttc ttgtgtgtgc actaggttag tctagagcag cacttcccaa aatgaggtcc 69960
cccagccagc agcatcagca taacctggaa attgttcaaa atgaagttcc agctaggtgc 70020
tgcagctcac gcctataatc ccagtacgtt gggaggccaa ggtgggagga tcacttgagc 70080
ccaggagtct agtctgtctg agaccagcct gggcaaaaaa gccagatatt gaaagaaaag 70140
aagagagaag aaaaggaaag aaaagaaaag aaaagaaaga aagggagaaa gagagagaga 70200
gaaagagaga gaaagagaaa gaaagaaaga aggaaggaag gaaggaagaa aaagaaagaa 70260
agaaagaaaa agaagaaacg caagttctca gccctcaccc aagactttgc agaccccgaa 70320
ttgctgggct gggctgggca tttgtgtgtg aactaccctc caggtggtca gaggcctggt 70380
gggaagttct ccaggcacct cccctgctct gagattgtat gtatccaaga acatttctct 70440

tctttttttct ccacacctat gtagcactat tgtttctttt tcagatacac atgctcactg 70500
tacacaataa agaaataact ttttttttttt ttttgagaca cagttgccat tctgtcaccc 70560
aggctggagt acagtggcac aatctcggct cactgcaacc tctacctcct ggattcaagt 70620
aattctcctg cctcagcctc cctagtagct gggattacag gcacatgcca ctatgctcag 70680
ctaatttttg tattattaat agaggcagag tgtcgccaag aaacaacctt tttgggccag 70740
gtgcggtggc tcacacctgt aatcccagca gtttgggaga ccgaggccgg cgaatcactt 70800
gaggtcagga gtttgagacc agcctggcca acatggtgaa accctgtctc tactaaaaat 70860
acaaaaatta gccaggcatg gtggcatgca cctgtaatcc cagctacttg ggaggctgag 70920
gcaggacaat cacttgaacc cgggaagcag aagttgcagt gagccaagat cgcaccactg 70980
cactccagct gcggtgacag tgagactctg tctcgaaaac aaaaacaaga acaaaaaacc 71040
ctttattgta taaaggtctt aataacctta atttcttctt ttttttttttt gagatgggat 71100
cttgctctgt tgcccagctg gagtgcagta gcatgatctc agctcactgc agcctctgcc 71160
tcctgagttc aagaattctc ctgcctcagc cccccaagta gctgggatta caggggtgtg 71220
ccaccacgcc tggctaattt ttgcattttt agtagagaca gggtttcacc atgttgggca 71280
ggctggtctt gaactcctga cctcaggtga tcgacctgcc ttagccttcc aaagtgctgg 71340
gattacaggc atgagccacc acacccggcc aataacctta atttcttaaa agtcattaag 71400
aaataacctt tatctggcag gagccctaag ccacagctct aataatccaa ccgttctcat 71460
ttttctgtct tccttttctag tcctttccta taggaatatg caaattaaaa accaattaag 71520
ttaatttttaa aaatccaatg catatcttga aaccatacag agaagaatct cggttcacta 71580
gggagatctc tgtaggcttc actcatcaaa ggtcaggcct gggtctccca cagcagtggg 71640
gccagctatg gagtttgcag ggctggtgca aaacaaaaat atgggcctct tgcacaaaat 71700
ttactaagaa tttcaaatgg tggtggcaga gccctgaacc ccgcttgatc acatgcctgt 71760
gccactgcgt ctgcggtgtt ctgaagttgt cctggaaagg gctctgacct ttgcccttcc 71820
atcttctgtg tgccatggct gtccagcctc caggttcatg gcctatcaca agcccctgaa 71880
aaactcacag gattacacag aagctctgcg ggcagctcga gagctggcag ccaacatcac 71940
tgctgacctg cggaaagtgc ctggaacaga cccggctttt gaggtcttcc cctacacgtg 72000
aggacctgag tggctgggct ggagggaggt ggggtatggt tgctggagac tggaggttag 72060
ggtggagggc ttgcaaggag ttgcatgaga tgaggaccag ttttaggtca ggaggctctg 72120
gctgcagcct tgggcctatt tcttaggctg gtttgtaccc caatataagc ctgcctgacc 72180
ctcagcattc tccttctgaa gtggggtgtc ccacccacca tgagggcccc agaggcctga 72240
gcctgtgacc atgctctgtg ctctggcagg atcaccaatg tgttttatga gcagtacctg 72300

accatcctcc ctgaggggct cttcatgctc agcctctgcc ttgtgcccac cttcgctgtc 72360
tcctgcctcc tgctgggcct ggacctgcgc tccggcctcc tcaacctgct ctccattgtc 72420
atgatcctcg tggacactgt cggcttcatg gccctgtggg gcatcagtta caatgctgtg 72480
tccctcatca acctggtctc ggtaacccag cagacacagg caccaggggg cctctggagg 72540
ggtggttggg gatccagcct catagaatac tcctagttct tttttgtttc ttttttaga 72600
ggcagggtct tgctctgttg ctcaggcttg agggcagtga catgatcaca gctcactgta 72660
gcctcgaacc cttgggctca agcgatcctc ctacctcagc ctccaaagta gccaggacta 72720
caggcacgtg ccactgcgtc cagctaatat tttaattttt gttgtagaga cagggtctca 72780
ctttgttgcc caggctggtc tcaaactcct gggctcaagt gatcctctca cctcggcctc 72840
ccaaagtgtt gggattatag gcatgagcca ctgcacccgg ccaaatactc ccagttctgt 72900
ctagaatcta gatgcctgcc ccacgctggt cctggtggag gcctcatctc cctagttcct 72960
tccccacctc tgcctttctt ggcttatgcc ccctctctgc ccataggcgg tgggcatgtc 73020
tgtggagttt gtgtcccaca ttacccgctc ctttgccatc agcaccaagc ccacctggct 73080
ggagagggcc aaagaggcca ccatctctat gggaagtgcg gtgagtggag aggagtgggc 73140
caccctgtgc cccactcgac accctgtgcc ctgcctgatg ccctgtgccc tgcctgatgc 73200
cctgtgccct gcctgacacc tggctctgaa ccccccaggt gtttgcaggt gtggccatga 73260
ccaacctgcc tggcatcctt gtcctgggcc tcgccaaggc ccagctcatt cagatcttct 73320
tcttccgcct caacctcctg atcactctgc tgggcctgct gcatggcttg gtcttcctgc 73380
ccgtcatcct cagctacgtg ggtgagtgcc caggcctgtt cctaccagac tgtcatgatt 73440
atgctgacga caacagtaac agtgcatgct caccacaaaa gctcaggaag tgcaaacgag 73500
ccatgggcag atgtcagaag ccaggactat gaccatgtgg caattctgtc ttggaagcta 73560
ctattattca tttaatgtgc tgtgaacatc tttttttgtc agctatgtat gtctcaaaca 73620
acgtttctgt ggccctgtac actgtggatc ttcactgcac tgctgttgga cttttaagca 73680
tgcccttcag caagaaatat attttacaca gagaggtgac atgcacgggc acacatagac 73740
atgcctgcct aaaacaaatg cttcactaaa taatattaat acttccttta tacatgtgaa 73800
gcattctgat attgctggtt ccattctatt attattatta atatttttttg gagacagggt 73860
cttgctctga cacccaggct ggagtgcagt agcatgatca cagctcactg ccaccttgac 73920
ttcccaggct caagtgatcc tcccacctca gcctcccgag tagctgggac cacaggtgca 73980
caccaccatg cccagctaat tttttatttt ttgtagagat ggggtctccc tatgttgccc 74040
aggctggtct caaactcctg agctcaagtg atccaccatg gccttccaca gtgctaggat 74100
tacaggtgtg agccactgcg cttggctttt attttacttt aaatttgtta tttatttttat 74160
tttactttac attattttat ttttatttttt tgagatggag tctcgctctg ttgcccaggc 74220

tggagtgcag tggtatgatc tcagctccct gcaacctctg cctcccaagt tcaagccatt 74280
ctcctgcttt agcctcccaa gtagctggga ttacaggtgc gcaccaccac gcctggccaa 74340
tttatttatt tatttttttat ttttagtaga gacggggttt caccatgttg ggcaggctgg 74400
tctcgaactc ctgacctcag gtgatccaac cgccaaggcc tcccaaagtg ctgggattac 74460
aggcgtgagc cactgtgccc agccctatca ttaatttgtt tttaattatt ttaattattt 74520
ttatttttat tatttttaga cagagtctct ctctgttgcc caggctggag tgcagtggcg 74580
caatctcagc tcactgcaac ctctgcctcc tgggttcaag cgattctcct gtctcagcct 74640
ctcgagtagc tgggatatcg gtgtatgcca ccatacctgg ctaatttttg tatttttatt 74700
ggagacaggt ttcaccatgt tggtcaggct ggtctcgaac tcctgtggcc tcaggtgatc 74760
catctgcctt ggcctcccaa agtgcaggga ttacaggcgt gggccaccgc acccggtctc 74820
attaatattt tgaaatgctg gccaggagtg gtggctcatg tttgtaatcc tagcactttg 74880
ggaggctgag gcacatggaa gctcaaattg agcctcccag gatgaaggtg tttctggctc 74940
tcagggtggg caagctggga ggagttcaat tttacctccc accagatggt aataatatta 75000
ttagaggaca tttatagagg ggtgtgtttg tgcatcaaca tatgtgtctg taattctctt 75060
actacccccg aggcaggtat tattatcctt cccattttac agatgaggga actgagacac 75120
ctgccccagg ttacagactt ggtcaaaggt agtaggggtt ggagcccaca cagctctgtg 75180
gttcctaacc atgtctcttg tggggactcc ctgaccctct tggaaggagt agagtgtgtg 75240
cgctgggggt ggtggatgag acataagaga ggggcaagga ggagcagtcg tggggtgtgc 75300
ttggacaaag gatatccagg gccttggagc tgcaggtggt ggctattcct tggaggttcc 75360
caaaatgctt gggggatgga gggaccagga catccctgaa gcttgggctg tgaacatagt 75420
gaccctggaa ggcacatggc acagatcccc cctgggaccc ttcctgccct gggtttgttg 75480
tacagaacca ggaatagctt ctcacctgtg tcccctgccc acctctctga ctgtggttct 75540
ctgtctctcc gcagggcctg acgttaaccc ggctctggca ctggagcaga agcgggctga 75600
ggaggcggtg gcagcagtca tggtggcctc ttgcccaaat caccctccc gagtctccac 75660
agctgacaac atctatgtca accacagctt tgaaggttct atcaaaggtg ctggtgccat 75720
cagcaacttc ttgcccaaca atgggcggca gttctgatac agccagaggc cctgtctagg 75780
ctctatggcc ctgaaccaaa gggttatggg gatcttcctt gtgactgccc cttgacacac 75840
gccctcctca aatcctaggg gaggccattc ccatgagact gcctgtcact ggaggatggc 75900
ctgctcttga ggtatccagg cagcaccact gatggctcct ctgctcccat agtgggtccc 75960
cagtttccaa gtcacctagg ccttgggcag tgcctcctcc tgggcctggg tctggaagtt 76020
ggcaggaaca gacacactcc atgtttgtcc cacactcact cactttccta ggagcccact 76080

tctcatccaa cttttccctt ctcagttcct ctctcgaaag tcttaattct gtgtcagtaa 76140
gtctttaaca cgtagcagtg tccctgagaa cacagacaat gaccactacc ctgggtgtga 76200
tatcacagga ggccagagag aggcaaaggc tcaggccaag agccaacgct gtgggaggcc 76260
ggtcggcagc cactccctcc agggcgcacc tgcaggtctg ccatccacgg ccttttctgg 76320
caagagaagg gcccaggaag gatgctctca taaggcccag gaaggatgct ctcataagca 76380
ccttggtcat ggattagccc ctcctggaaa atggtgttgg gtttggtctc cagctccaat 76440
acttattaag gctgttgctg ccagtcaagg ccacccagga gtctgaaggc tgggagctct 76500
tggggctggg ctggtcctcc catcttcacc tcgggcctgg atcccaggcc tcaaaccagc 76560
ccaacccgag cttttggaca gctctccaga agcatgaact gcagtggaga tgaagatcct 76620
ggctctgtgc tgtgcacata ggtgtttaat aaacatttgt tggcagaaat ggtgttttat 76680
gtcacatgtc ctaccctggc ttcctcctct cggtttaaga taatttttgt gaatgacaca 76740
aataatacat gtgtgggaga gtgatttgtg gagatactag tctgtgtttt gttctatttc 76800
tcctccctct tttcaagaaa gtagccaggc cattgtgtgc tcatgcctta caagggcctt 76860
tgaggagtgg gagtaatttc tcttcaaact gggagggcac agagcctgag agtcagtcag 76920
gagtaggatg tgcagcccct ccttttctgg aagagactgt gaagtaggca acacctggag 76980
gagctacagg agaaccacgg tgcattcaag gagggaagaa cccaccgtac aaacaaccag 77040
ctcccaggag ggccccaggc cagggcagtg ggtggaaatg tcaaggaaca ttccagatcc 77100
cctcgagtct ttctgcccca tgctgggtcc agcccttgtt tggctgaggg gctgctgttg 77160
ctttgaggct cagagggact gtcagcatgt aaagggaaga caagcaaaaa ggggtggaaa 77220
ggagctggcg tttctggagc ctactatcta cttttgggtc ctcataagag ccccatgtgc 77280
cagcatcatt agcccacctt tgggagggtt gctggctgac catgatggac aggaggtttg 77340
gtgaagggac agctacgagg gaatagaggc tgaggagaaa tcgcacaatt caccctgtta 77400
aaaactccac aggtgcagaa taaacagata gatttgagga acaaaatagc ttttgacagc 77460
agacatttca aatcagagga aagggtagat ccttcagtaa acggtgtgag agtagtgagc 77520
aaattatttg gatcaaaata aagttatatc tatacttcac acaatacaca aaataaaagt 77580
acagacagat taaagcacta aacacaaaaa tgaaactata caactatcgg aaggaaacac 77640
agaagagtat gttataatct tggaggggga aaagtttcct aagcacaaag tccagaagcc 77700
ataaaggtaa acactaaggt atgaccatat aataatggaa aacatctgaa aacacacaaa 77760
aaattaaaga aagttgaaag acacatatga gctcagaaaa atagttgcaa catatttaac 77820
agcaaataaa atcaagaaaa cacaaagagt gccaatagtg ctcctgcaaa catggtgaac 77880
actcctaaaa cccactggac tttctgtaag aagtgtggga agcaccagcc ccacagagtg 77940
acacaggaca catttccctg tatgcctagg gaaagccatg ttatgacagg aagcagaggg 78000

gctatggtgg ggagactaag ccaattttcc agaaaaaggc taaaactaca aagaagattg 78060
tgctaagttt tgagtgcatg aagcccaact gcagatctaa gagaatgctg gctattaaga 78120
gatacaagca ttttgaactg ggaggaggta agaagagcaa gggccaagtg atccagttct 78180
aagtgtcatc ttttgtttta ttatgaagac aataaaatat tgagtttatg tttaaaaaaa 78240
aaaagaatat acaaagagag tccaggtacg gtggctcatg cctgtaatcc cagcactttg 78300
ggaggctgag gcaggagaat tgcttgaggc caggagttca agaccagcct aggcaacata 78360
gcgagatact gtctctacaa aaagtttaaa agttagccag gctagctatt tggaaggctg 78420
aggtgggagg attgtttcag ctcgagtttg aggctgcagt gagctatgat ggcaccactg 78480
tactccagcc tgagtgaaag agtgagcttc tgtctcaaca aaaaaaaaaa aaaaaaagaa 78540
tatacaaaga gaggaaggag tgcaggggggg aggtctgggt tatgtggcta accttcccat 78600
tagaaacaag acattctagc taaaataaat cttagccgtg tgtgtgtgtg tatgtgtctg 78660
tgtgtgtgta tgatgcatac aagtttaggg tgttttaacc ttcttgataa attgagactt 78720
ttatagtttg aaatgactat aaaaatatcc ctttttatct ctagtattta tttttgtctg 78780
tttaagagat ggggttctca ctttgttgcc caggctggtc ttgaatactt ggcctcaagg 78840
gatcctccta cctcagcctc ccaagtacct ggaattacag gtatgagcca ccatgccagt 78900
cctatctgta gtatttgttc aactgtataa tgttattata cacacacaca cacacacaca 78960
cacacacaca cagacacaca cacacatata aaataacata cggttgaaca aattttatac 79020
ttaatagtca aacattgaaa ccctttcccc tgagattggg aatgagacaa agttgcccac 79080
ttttacccaa cattgcactg gaggtcttag ccattgtaat aaggcaagaa aaagaaacta 79140
agtttataag gattagaaat aaataaaatt gacatcattc acagataaca taaatatgta 79200
taaaaaagat tcagtctggg tgcagtggct catgcctgta accccagcaa tttctgaggc 79260
caaggcagga ggatcacttg aggccaggag ttcaagacat agcaagaccc cacctctaca 79320
aaaaaaaatt ttttttaaag atccaaaaga atctatatat aaactattgg aattactcta 79380
acaaaaggtg gtcaagaaaa ctatgaaaaa taataacttt gtattttaat ttgtataata 79440
ttgagagaaa ttaactgtca aaagaaatgg aggaatatac catgaattga gggctctata 79500
ctacagagat gtcaattctc ttcaaattaa ttactagttt cactgtaatt tcaataataa 79560
ccccagaaaa ttttttgtgg aaactgataa gctgattcaa aaattcatat agaaccacaa 79620
aagatgaaaa ttcacgaaag caatcttgaa gaaaaacaaa gtcagagaac ttacactact 79680
agaaatcaag ataatataaa tatatagaaa taaagatagt gagattttgg cacaaggaag 79740
aacaaataga aaaatggaaa gaatagaaag tccagaaaca gatgataccc acaaggacac 79800
atgatttatg atggaggagg catgcagagc attgggtaaa ggaggttttt caatgtagga 79860

tgctgaccta gttgggtatc cacacagaaa gaaatgaatc atgaccctct cccccaagat 79920
acacaaaaat cagttcctga tagattgtca atctaaatgt gaaagataaa atgatagagt 79980
tctaaaaggt aacataaaag agtatcccca agactgaaat aggaaaaact tttcttagga 80040
aacaaaagcc ttacttatag agaaaaagat tgataaattg aactgtattg gaataaaaaa 80100
aaacttctgt tcttcaaaag acatccttag gaaagataaa attcaaacca tagagaggaa 80160
aagatatttg cacatatctg aaatacacac atatctgaga aagggcctgt gcttagaatg 80220
cataaaaaat ctcctacaac tcagcaagaa aaagacagac aaccaaaaga aaagctaggc 80280
tggctactca aataagcaaa tggccaatac aagttcctca attttgtcag tcaccagagc 80340
aaggctgagt aaaagcacag tgagagttct tcctcttctc ttccctcaca atttggccta 80400
caggccatgg ggtaaggtgg ggccaggcag cacatgtggg gtgtcagaat ccaggtggtg 80460
tggggagcgt ttccacattg gatctgaggg aggagaggag ggcattccac acagaatagg 80520
aactacatag gcccagtatg gggctaagat gtcagaactg agctctgatg tgcctttctc 80580
catgagcaga gggactggat gctggagatg gagggtggag gaaaggttca gagccatcta 80640
gagatggcaa ttcagaggaa atgggagggc agatagtctc actcttcaca gtgaggcaga 80700
gtttccaagc tggttttgtc actcctttgc tgggcctctt tgggtaacat atttgactta 80760
tctgggcttt agtttctttt ttgctttttt ttttttttga gacagagtct cactctgttg 80820
gccaggctgg agtgcactgg tgtgatctta gctcactgca acctctgcct cccgggtttg 80880
agtgattctc ctgcctcagc ctcccgagta gctgcaacta caggcgcctg ccaccatgcc 80940
tggctaattt ttgtatacag atagggtttt gccatgttga ccaggctggt cttgaactcc 81000
tgacccgagg tgatatgcct gcctcagcct cccaaattgc taggattaca ggtgtgagcc 81060
accacacctg gcatgggttt ggtttcttta cctgtaaaaa ctgggatagt ttagctgggc 81120
acagtgatgc taattgttgt cccagctact tgagaggctg agatgggagg atcacttgag 81180
cctaagaatc gcaggtcagc ctgggcaaca tagcaatacc ccatctgtga aaaaaaaaat 81240
tagtggctga gcacagtggc tcactccagc aatcccagaa ctttgggagg ccaaggtggg 81300
aagattactt gagcccagga gtttgaaact ggtctgggaa acacacagag accacaatct 81360
ctgcattaaa aaaaaaatta gctgggttgg tggcactcac ctgtggtccc agctacttgg 81420
gagggtgagg tgggaggata atttgatccc aggaagtgga ggctgcagag agctgtgatc 81480
atgccactgc actccagcct gggtcacaga gtgagaccct gtctcaaaaa aaaaaaaaaa 81540
aattaggaaa atttgccctg actccccacg ttttttttaa aggatgaaat gagatattat 81600
atgtgaaagc atctagtact tgtgacatag taggtgctta aaaagtgttt ccacttcact 81660
tctgcctaaa acccagttca gttcctgagt tccagatatc taactgtgat gagaagagac 81720
gcagccagag gtacctcaaa gatagcaaca ccccctccg ccccgatacc tgatgtactg 81780

aagtcagaaa tttaaaaaaa aaccttgttc ttccttcagt tttaagttca gtatactgat 81840
gaactatcgg tcacatttga cgatttactt taaaaataaa caggcttcca aattaaccta 81900
cttatatggt ttgtctgtgt cgccacccaa atctcatctt gaattcccac ccgttgtggg 81960
agggacctgg tgggaggtaa ttaaatcatg ggggcaagtc tttcctgtgc tgttctcgtg 82020
atagtgaata agtctcaaga gatctgatgg ttttaaaaag aggagttccc ttgcacaagc 82080
tctctctctt tgcctgctgc catccatgta ggatgtgact tgctcttcct tgccttccac 82140
catgattgtg aggcttcccc agccacatgg aactgtaact ccaattaaac ctctttctct 82200
tgtaaattgc ccagtctagg ctatgtcttt atcagcagtg tgaaaacaga ctaatatact 82260
taccttggaa aggccttgtg atccatggtg acatcttgtc cctaaggaaa gcatcttacc 82320
atgagttcct caaattgttg atgtactgat taatgtgtaa ccctctgaca ctgggaagaa 82380
cactgattta tttctgaatc ataaagtttt attgattgtc ttgcatgtag acattttagc 82440
ttgtatgttg caatctgtat ccaacaattg taacctctgt attgtaccct caaatgaaag 82500
aggaaaaaac tcttgtatga ggagtcccct cccttctcct aaactttcct ataaaagcct 82560
tctaccttgt aacagactgg aacattccta acattgttgg tgtgtttcct aagcggattc 82620
tcacatttgg cttcaaataa accttgatca aattagtgct gcctcaacag ccttaatttc 82680
aatcaatagt acaagcctct gtttttctat ttaatcacta ctttaaaggt aacctttgga 82740
aaatatttag gctctttaca aatttaatta attgaacata ttttaactgc atttataaag 82800
gtaatagtct ccatttttctt cctaaatact ctgcataaga aacaaaatct tcccatatac 82860
ttaactcttt taaacctaat aaattaaatt tatggaatat cattaatata aagtttttat 82920
agatgttgta acactgcaca tagatttagc aacatttcaa tttacaatct taagcttata 82980
tgaaatacca ttttaaattg gaattataca attcttacac taatagacca aatactttaa 83040
atgttacaag catataaaat acgaaatata caaaaatttc cccccatcac acaaatattc 83100
ttactaaggt tttgcttctt tgaaaccttt ctatacacat tgtattagtc tgttttcatg 83160
ctgctaataa agacataccc cagactgggt aatttataaa ggaaagagtt ttaattgact 83220
tatagttcag catggctggg gaggcctcag gaaacttaca atcatggcag agggggaagc 83280
aaacatgccc ttcttcatat ggcaccagtg gagagaagaa tcagtgccca gtgaaagggg 83340
aagcccctta taaaaccagc agatctcgtg agaactaaat cactaccaca agaacaggat 83400
gggggaaacc gctctcatga ttcaacgatc tccacctggt ccctcccaca acacatgggg 83460
attatgcaaa ctgcaagtca agatgagatt tgggtgggga cacagtcaaa acctatcaac 83520
ctaacatcct tttcctctcc ccttccttcc ttcctccctt ccttccttcc ttccttcctt 83580
ccttccttct ttccttccct ccctccctcc ctccccctct ctctcttttt ttctttttct 83640

tttctttctt tctctctctc tctctccctc cctccctccc aggctggaat gcagtggtgc 83700
gatctcggct cactgcaacc tctgcctccc agggttaagc tatcctccca cctcagcctc 83760
ctgagtaact ggtgggacta caggcgtgtg acaccacacc cagctgatgt ttttgtattt 83820
ttagtggaga tggggtttca gtatgttgtc caggctgtcc atacccattt ttaagtgagt 83880
tataaatggg gttcaaaggt catactcccc ttgaggaaga caatcatcat ctcagataac 83940
caaggttgcc tatgcagtaa ggaagaagta agtcatcatt ccgggtaact aaatttacct 84000
aagaccaaag acatcagctg agagtgagac ctggagtctc aggcatcggg agtagttatc 84060
tcactgctaa ctaagtttac atggtgagtc aaaagaccca gaatacccaa cacaatattg 84120
aaggaaaaca aagtcagagg actaacacta tctgacttct agacttacta taaagttata 84180
gtaatgaaga cagtgaaaga actggtaaag aacagataaa taaatcactg taacagaata 84240
tagagtctag aaatagaccc aaataaatat agtgaagcaa aggtagactt ttttttttt 84300
ttttttttt tgagacagag tctctctctg tcacccaggc tggagtgcac tggtatgatc 84360
ttggttcaat gggacctata cctttaccat gagaatcact gggttcaagt gattctcatg 84420
cctcagtgtc ctgcatagct gggactaaag gcctgcaaac atgcctggct aatttttgca 84480
tttttagtag agatggggtt tcaccatgtt ggctaggctg gtctcaaagt cctgacctca 84540
ggtgatccac ccgccttggc ttcccaaagt gctgggatta caggtgtgag ccaccatacc 84600
cagccaaagg gcagtctttc caacaaatga tacagataca actggacatc tatgtgcaaa 84660
aacataaatt tagacacaga ctttgcaccc ttcacaaaaa ctaactgaaa atggatcata 84720
gacctcaatg taaaattcaa aactataaaa ctcctaaaag acaacatagg gtaaaaccta 84780
gatgaccttg ggtgtagcga ccttttgata caacaccaaa gacataatcc atgaaataaa 84840
taactgataa actgtaatta ataaattttt tttagcagta atagaatgat gagtgttatt 84900
tcattaaaat ttaaaacttc tgctctgcaa aagacaatgt caagaagaag aagacaatgg 84960
ccaagtgcgg tggcttatgc ctttaatcct agcactttgg aaggccaagg cgggtggatc 85020
acttgaggcc aggagtttga gaccagcctg gctaacatgg tgaaaacctg tctctactaa 85080
aaatagaaaa attagctggg cgcagtggtg cacacctgta atcccagcta cttgacaggc 85140
tgatgcacaa gaatcgcttg aacccaggag gcagaggttg cagtgagctg aaattgtgcc 85200
actgtactcc agcttgggca acagagcgag actctgtctc aaaaaatata taaataaata 85260
aaatttaaaa aggatgagaa gacaagccac tgcctgggag aagatattag cgaaagacac 85320
atctgtgctg gcttcagcag cacacatact aaaattacaa tggtacagag aagattacca 85380
tggcctgtgc acaaggatga catgcacatt tgtgaagtgc ttcagaatat aaaaaagaaa 85440
aagatctatc cgataaagaa cttttattta aaatctaaat ggactctcca atacaataat 85500
aagaaaacaa ataactcaat taaaaactta gccttaccaa agaagatgta cagatggcaa 85560

acaagcatat gaaaagatgc tacatgtcat atatcatcag ggaaatgata attaaaacaa 85620
caatgtgata ctgctacaca tctattagaa tgtccaaaat ctggaacact gacaacatca 85680
aatgctggta gggatgtgga gaagcagcaa ctctccttca ttactgatag gaatgcaaaa 85740
tggtacagcc actttggaag acagttcttc agtttcttct aaaactaaat atatcttacc 85800
atatgatcca gcaatcacat ttcttggtat gtacccaatg gagttgaaaa cttatgtcta 85860
cacaaaaacc aacacatggg tgttcatggc agccttattc ataattgtca aaacttggaa 85920
gtaaccaaga tgtccttcag taggtgaatg ggttaatccc cacaatggaa tattattcag 85980
cattaaaaac aaatgagcta tcaagctaag ctatgaaaag acatggaggg gccgggcacg 86040
gtggctcaag tctgaaatcc cagcactttg ggaggccgag gtgggcagat cacaaggtca 86100
ggagtttgag accagcccgg ccaatatggt gaaaccctgt ctctactaaa aatacaaaaa 86160
ttcgccgggt gtggtggcag gcgcctatag tcccagctac tcagatggct gaggtaggag 86220
aggagattca cttgaatctg ggaggcagag gttgcagtga gccgagatca caccattgca 86280
ctccagcctg ggcaacaaga gcgaaactcc atctcaaaaa aaaaaaaaaa aaaaaagaaa 86340
aagaaaaaga aagaaaagaa aagaagtgga ggaacttcaa atgtatacta ctaagtggaa 86400
aaagcaaatc taaaaagtct acatctgtct gattccaact atatgacatt ctgtaaaagg 86460
caaagctata aacacaataa aatgatcggt agtttctagg gtttggggtt aggggagttg 86520
aatgggcaga gcacaaaaga tttttaggcc agggaaacca ctctatatga tattataatc 86580
atggatgcat gtcattatac atttgtccaa atccatagaa tgtacaacac cagagtgagc 86640
cctaatgtaa actaaggatt ctgggtgata ttgtaacaaa tgcaccatta attgtaacaa 86700
atgtatcatt ttgtaccttc tgatggggaa tgttgagaat gagagaggct atgcatgtgt 86760
ggaggcagga gtgggtatat gggatatctc tgtatcttcc tctcaatttt gctgtgaacc 86820
tataactacc taaaaaagtc ttttagaaag cccagtagtt ttttgcttct ctttatgggt 86880
tggtttcctt ctctcaagtg aaaaatgggc ttcctccatg tagcagatga tatggcttct 86940
ctcatcccag agaagagagt tctttcttgt caattacagc cagaaaaatc tccaagaagg 87000
atttagatgg tcctagtttg ctccctccca tccctcttcc tttggatctc agatcagaag 87060
tgacttctac tgggatgctg ccctgttacc ccagtcttgg tcgggtccct gttatgtgct 87120
cccactatac catatccttc tccttcctag tcttcatcac agtttgaaga tgaaaattca 87180
ttggtggggt tacatggctc ccccatgtct gattcctcct ctaaactgta agctataggg 87240
ggcaatgact ttatttttt gcttaccatt gtgtttctag cacctagcat ctggcacata 87300
ggcacacaat aaatatccat taaataaatg actgaaataa acagagggct cttttgctct 87360
gattactctg aagagcaatt attacatagc agtgacagct tagtgtattc tcagaaaata 87420

ttcttttgtt ttaaaaccac ttatttttct ggccaggcat ggtggctcac gcctgttatc 87480
tcagcacttt gggaggccga ggtaggcgga tcacaaggtc aggagatcga gaccatcctg 87540
gataacatgg tgaaaccctg tctctactaa aaatacaaaa aaatgagccg ggcttggtgg 87600
cgggcgcctg tagtcccagc tactagggag gctgaggcag gagaatggcg tgaacccagg 87660
aggtggaggt tgccatgagc caagatcgca ccgctgaact tcagcttggg cgacagagcg 87720
agattccatc tcaaaaaaaa aaaatttttt tttctgataa taaacacaac agactgggca 87780
cagtggcgca tacctgtaat cctggtacat tgggaggcca aggtgggagg atcacttgag 87840
tccaggagtt caagaccagc ctgggcaaca ttgtgagaca tcatctctat ttaaaaacaa 87900
acaaacaaac aaacaaacaa acaaacaaac actccttaaa tccccacaca cttatgacag 87960
aataattgta agacaaagaa aagtacagtt aagaaaacaa aaaacaaaaa ttacttatat 88020
ctgtaaccc 88029

<210> 21
<211> 5092
<212> DNA
<213> Homo sapiens

<400> 21
cttggctgtt cctgaggcct ggcctggctc cccgctgacc ccttcccaga cctgggatgg 60
cggaggccgg cctgaggggc tggctgctgt gggccctgct cctgcgcttg gcccagagtg 120
agccttacac aaccatccac cagcctggct actgcgcctt ctatgacgaa tgtgggaaga 180
acccagagct gtctggaagc ctcatgacac tctccaacgt gtcctgcctg tccaacacgc 240
cggcccgcaa gatcacaggt gatcacctga tcctattaca gaagatctgc ccccgcctct 300
acaccggccc caacacccaa gcctgctgct ccgccaagca gctggtatca ctggaagcga 360
gtctgtcgat caccaaggcc ctcctcaccc gctgcccagc ctgctctgac aattttgtga 420
acctgcactg ccacaacacg tgcagcccca atcagagcct cttcatcaat gtgacccgcg 480
tggcccagct aggggctgga caactcccag ctgtggtggc ctatgaggcc ttctaccagc 540
atagctttgc cgagcagagc tatgactcct gcagccgtgt gcgcgtccct gcagctgcca 600
cgctggctgt gggcaccatg tgtggcgtgt atggctctgc cctttgcaat gcccagcgct 660
ggctcaactt ccagggagac acaggcaatg gtctggcccc actggacatc accttccacc 720
tcttggagcc tggccaggcc gtggggagtg ggattcagcc tctgaatgag ggggttgcac 780
gttgcaatga gtcccaaggt gacgacgtgg cgacctgctc ctgccaagac tgtgctgcat 840
cctgtcctgc catagcccgc ccccaggccc tcgactccac cttctacctg ggccagatgc 900
cgggcagtct ggtcctcatc atcatcctct gctctgtctt cgctgtggtc accatcctgc 960
ttgtgggatt ccgtgtggcc cccgccaggg acaaaagcaa gatggtggac cccaagaagg 1020

gcaccagcct ctctgacaag ctcagcttct ccacccacac cctccttggc cagttcttcc 1080
agggctgggg cacgtgggtg gcttcgtggc ctctgaccat cttggtgcta tctgtcatcc 1140
cggtggtggc cttggcagcg ggcctggtct ttacagaact cactacggac cccgtggagc 1200
tgtggtcggc ccccaacagc caagcccgga gtgagaaagc tttccatgac cagcatttcg 1260
gccccttctt ccgaaccaac caggtgatcc tgacggctcc taaccggtcc agctacaggt 1320
atgactctct gctgctgggg cccaagaact tcagcggaat cctggacctg gacttgctgc 1380
tggagctgct agagctgcag gagaggctgc ggcacctcca ggtatggtcg cccgaagcac 1440
agcgcaacat ctccctgcag gacatctgct acgcccccct caatccggac aataccagtc 1500
tctacgactg ctgcatcaac agcctcctgc agtatttcca gaacaaccgc acgctcctgc 1560
tgctcacagc caaccagaca ctgatggggc agacctccca agtcgactgg aaggaccatt 1620
ttctgtactg tgccaatgcc ccgctcacct tcaaggatgg cacagccctg gccctgagct 1680
gcatggctga ctacggggcc cctgtcttcc ccttccttgc cattgggggg tacaaaggaa 1740
aggactattc tgaggcagag gccctgatca tgacgttctc cctcaacaat taccctgccg 1800
gggacccccg tctggcccag gccaagctgt gggaggaggc cttcttagag gaaatgcgag 1860
ccttccagcg tcggatggct ggcatgttcc aggtcacgtt catggctgag cgctctctgg 1920
aagacgagat caatcgcacc acagctgaag acctgcccat ctttgccacc agctacattg 1980
tcatattcct gtacatctct ctggccctgg gcagctattc cagctggagc cgagtgatgg 2040
tggactccaa ggccacgctg ggcctcggcg gggtggccgt ggtcctggga gcagtcatgg 2100
ctgccatggg cttcttctcc tacttgggta tccgctcctc cctggtcatc ctgcaagtgg 2160
ttcctttcct ggtgctgtcc gtgggggctg ataacatctt catctttgtt ctcgagtacc 2220
agaggctgcc ccggaggcct ggggagccac gagaggtcca cattgggcga gccctaggca 2280
gggtggctcc cagcatgctg ttgtgcagcc tctctgaggc catctgcttc ttcctagggg 2340
ccctgacccc catgccagct gtgcggacct ttgccctgac ctctggcctt gcagtgatcc 2400
ttgacttcct cctgcagatg tcagcctttg tggccctgct ctccctggac agcaagaggc 2460
aggaggcctc ccggttggac gtctgctgct gtgtcaagcc ccaggagctg cccccgcctg 2520
gccagggaga ggggctcctg cttggcttct tccaaaaggc ttatgccccc ttcctgctgc 2580
actggatcac tcgaggtgtt gtgctgctgc tgtttctcgc cctgttcgga gtgagcctct 2640
actccatgtg ccacatcagc gtgggactgg accaggagct ggccctgccc aaggactcgt 2700
acctgcttga ctatttcctc tttctgaacc gctacttcga ggtgggggcc ccggtgtact 2760
ttgttaccac cttgggctac aacttctcca gcgaggctgg gatgaatgcc atctgctcca 2820
gtgcaggctg caacaacttc tccttcaccc agaagatcca gtatgccaca gagttccctg 2880
agcagtctta cctggccatc cctgcctcct cctgggtgga tgacttcatt gactggctga 2940

ccccgtcctc ctgctgccgc ctttatatat ctggccccaa taaggacaag ttctgcccct 3000
cgaccgtcaa ctctctgaac tgcctaaaga actgcatgag catcacgatg ggctctgtga 3060
ggccctcggt ggagcagttc cataagtatc ttccctggtt cctgaacgac cggcccaaca 3120
tcaaatgtcc caaaggcggc ctggcagcat acagcacctc tgtgaacttg acttcagatg 3180
gccaggtttt agacacagtt gccattctgt cacccaggct ggagtacagt ggcacaatct 3240
cggctcactg caacctctac ctcctggatt cagcctccag gttcatggcc tatcacaagc 3300
ccctgaaaaa ctcacaggat tacacagaag ctctgcgggc agctcgagag ctggcagcca 3360
acatcactgc tgacctgcgg aaagtgcctg gaacagaccc ggcttttgag gtcttccccct 3420
acacgatcac caatgtgttt tatgagcagt acctgaccat cctccctgag gggctcttca 3480
tgctcagcct ctgccttgtg cccaccttcg ctgtctcctg cctcctgctg ggcctggacc 3540
tgcgctccgg cctcctcaac ctgctctcca ttgtcatgat cctcgtggac actgtcggct 3600
tcatggccct gtggggcatc agttacaatg ctgtgtccct catcaacctg gtctcggcgg 3660
tgggcatgtc tgtggagttt gtgtcccaca ttacccgctc ctttgccatc agcaccaagc 3720
ccacctggct ggagagggcc aaagaggcca ccatctctat gggaagtgcg gtgtttgcag 3780
gtgtggccat gaccaacctg cctggcatcc ttgtcctggg cctcgccaag gcccagctca 3840
ttcagatctt cttcttccgc ctcaacctcc tgatcactct gctgggcctg ctgcatggct 3900
tggtcttcct gcccgtcatc ctcagctacg tggggcctga cgttaacccg gctctggcac 3960
tggagcagaa gcgggctgag gaggcggtgg cagcagtcat ggtggcctct tgcccaaatc 4020
acccctcccg agtctccaca gctgacaaca tctatgtcaa ccacagcttt gaaggttcta 4080
tcaaaggtgc tggtgccatc agcaacttct tgcccaacaa tgggcggcag ttctgataca 4140
gccagaggcc ctgtctaggc tctatggccc tgaaccaaag ggttatgggg atcttccttg 4200
tgactgcccc ttgacacacg ccctcctcaa atcctagggg aggccattcc catgagactg 4260
cctgtcactg gaggatggcc tgctcttgag gtatccaggc agcaccactg atggctcctc 4320
tgctcccata gtgggtcccc agtttccaag tcacctaggc cttgggcagt gcctcctcct 4380
gggcctgggt ctggaagttg gcaggaacag acacactcca tgtttgtccc acactcactc 4440
actttcctag gagcccactt ctcatccaac ttttcccttc tcagttcctc tctcgaaagt 4500
cttaattctg tgtcagtaag tctttaacac gtagcagtgt ccctgagaac acagacaatg 4560
accactaccc tgggtgtgat atcacaggag gccagagaga ggcaaaggct caggccaaga 4620
gccaacgctg tgggaggccg gtcggcagcc actccctcca gggcgcacct gcaggtctgc 4680
catccacggc cttttctggc aagagaaggg cccaggaagg atgctctcat aaggcccagg 4740
aaggatgctc tcataagcac cttggtcatg gattagcccc tcctggaaaa tggtgttggg 4800

tttggtctcc agctccaata cttattaagg ctgttgctgc cagtcaaggc cacccaggag 4860

tctgaaggct gggagctctt ggggctgggc tggtcctccc atcttcacct cgggcctgga 4920

tcccaggcct caaaccagcc caacccgagc ttttggacag ctctccagaa gcatgaactg 4980

cagtggagat gaagatcctg gctctgtgct gtgcacatag gtgtttaata aacatttgtt 5040

ggcagaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa 5092

<210> 22
<211> 1359
<212> PRT
<213> Homo sapiens

<400> 22

Met Ala Glu Ala Gly Leu Arg Gly Trp Leu Leu Trp Ala Leu Leu Leu
1 5 10 15

Arg Leu Ala Gln Ser Glu Pro Tyr Thr Thr Ile His Gln Pro Gly Tyr
20 25 30

Cys Ala Phe Tyr Asp Glu Cys Gly Lys Asn Pro Glu Leu Ser Gly Ser
35 40 45

Leu Met Thr Leu Ser Asn Val Ser Cys Leu Ser Asn Thr Pro Ala Arg
50 55 60

Lys Ile Thr Gly Asp His Leu Ile Leu Leu Gln Lys Ile Cys Pro Arg
65 70 75 80

Leu Tyr Thr Gly Pro Asn Thr Gln Ala Cys Cys Ser Ala Lys Gln Leu
85 90 95

Val Ser Leu Glu Ala Ser Leu Ser Ile Thr Lys Ala Leu Leu Thr Arg
100 105 110

Cys Pro Ala Cys Ser Asp Asn Phe Val Asn Leu His Cys His Asn Thr
115 120 125

Cys Ser Pro Asn Gln Ser Leu Phe Ile Asn Val Thr Arg Val Ala Gln
130 135 140

Leu Gly Ala Gly Gln Leu Pro Ala Val Val Ala Tyr Glu Ala Phe Tyr
145 150 155 160

Gln His Ser Phe Ala Glu Gln Ser Tyr Asp Ser Cys Ser Arg Val Arg
165 170 175

Val Pro Ala Ala Ala Thr Leu Ala Val Gly Thr Met Cys Gly Val Tyr

180 185 190

Gly Ser Ala Leu Cys Asn Ala Gln Arg Trp Leu Asn Phe Gln Gly Asp
195 200 205

Thr Gly Asn Gly Leu Ala Pro Leu Asp Ile Thr Phe His Leu Leu Glu
210 215 220

Pro Gly Gln Ala Val Gly Ser Gly Ile Gln Pro Leu Asn Glu Gly Val
225 230 235 240

Ala Arg Cys Asn Glu Ser Gln Gly Asp Asp Val Ala Thr Cys Ser Cys
245 250 255

Gln Asp Cys Ala Ala Ser Cys Pro Ala Ile Ala Arg Pro Gln Ala Leu
260 265 270

Asp Ser Thr Phe Tyr Leu Gly Gln Met Pro Gly Ser Leu Val Leu Ile
275 280 285

Ile Ile Leu Cys Ser Val Phe Ala Val Val Thr Ile Leu Leu Val Gly
290 295 300

Phe Arg Val Ala Pro Ala Arg Asp Lys Ser Lys Met Val Asp Pro Lys
305 310 315 320

Lys Gly Thr Ser Leu Ser Asp Lys Leu Ser Phe Ser Thr His Thr Leu
325 330 335

Leu Gly Gln Phe Phe Gln Gly Trp Gly Thr Trp Val Ala Ser Trp Pro
340 345 350

Leu Thr Ile Leu Val Leu Ser Val Ile Pro Val Val Ala Leu Ala Ala
355 360 365

Gly Leu Val Phe Thr Glu Leu Thr Thr Asp Pro Val Glu Leu Trp Ser
370 375 380

Ala Pro Asn Ser Gln Ala Arg Ser Glu Lys Ala Phe His Asp Gln His
385 390 395 400

Phe Gly Pro Phe Phe Arg Thr Asn Gln Val Ile Leu Thr Ala Pro Asn
405 410 415

Arg Ser Ser Tyr Arg Tyr Asp Ser Leu Leu Leu Gly Pro Lys Asn Phe
420 425 430

Ser Gly Ile Leu Asp Leu Asp Leu Leu Leu Glu Leu Leu Glu Leu Gln
435 440 445

Glu Arg Leu Arg His Leu Gln Val Trp Ser Pro Glu Ala Gln Arg Asn
450 455 460

Ile Ser Leu Gln Asp Ile Cys Tyr Ala Pro Leu Asn Pro Asp Asn Thr
465 470 475 480

Ser Leu Tyr Asp Cys Cys Ile Asn Ser Leu Leu Gln Tyr Phe Gln Asn
485 490 495

Asn Arg Thr Leu Leu Leu Leu Thr Ala Asn Gln Thr Leu Met Gly Gln
500 505 510

Thr Ser Gln Val Asp Trp Lys Asp His Phe Leu Tyr Cys Ala Asn Ala
515 520 525

Pro Leu Thr Phe Lys Asp Gly Thr Ala Leu Ala Leu Ser Cys Met Ala
530 535 540

Asp Tyr Gly Ala Pro Val Phe Pro Phe Leu Ala Ile Gly Gly Tyr Lys
545 550 555 560

Gly Lys Asp Tyr Ser Glu Ala Glu Ala Leu Ile Met Thr Phe Ser Leu
565 570 575

Asn Asn Tyr Pro Ala Gly Asp Pro Arg Leu Ala Gln Ala Lys Leu Trp
580 585 590

Glu Glu Ala Phe Leu Glu Glu Met Arg Ala Phe Gln Arg Arg Met Ala
595 600 605

Gly Met Phe Gln Val Thr Phe Met Ala Glu Arg Ser Leu Glu Asp Glu
610 615 620

Ile Asn Arg Thr Thr Ala Glu Asp Leu Pro Ile Phe Ala Thr Ser Tyr
625 630 635 640

Ile Val Ile Phe Leu Tyr Ile Ser Leu Ala Leu Gly Ser Tyr Ser Ser
645 650 655

Trp Ser Arg Val Met Val Asp Ser Lys Ala Thr Leu Gly Leu Gly Gly
660 665 670

Val Ala Val Val Leu Gly Ala Val Met Ala Ala Met Gly Phe Phe Ser
675 680 685

Tyr Leu Gly Ile Arg Ser Ser Leu Val Ile Leu Gln Val Val Pro Phe
690 695 700

Leu Val Leu Ser Val Gly Ala Asp Asn Ile Phe Ile Phe Val Leu Glu
705 710 715 720

Tyr Gln Arg Leu Pro Arg Arg Pro Gly Glu Pro Arg Glu Val His Ile
725 730 735

Gly Arg Ala Leu Gly Arg Val Ala Pro Ser Met Leu Leu Cys Ser Leu
740 745 750

Ser Glu Ala Ile Cys Phe Phe Leu Gly Ala Leu Thr Pro Met Pro Ala
755 760 765

Val Arg Thr Phe Ala Leu Thr Ser Gly Leu Ala Val Ile Leu Asp Phe
770 775 780

Leu Leu Gln Met Ser Ala Phe Val Ala Leu Leu Ser Leu Asp Ser Lys
785 790 795 800

Arg Gln Glu Ala Ser Arg Leu Asp Val Cys Cys Cys Val Lys Pro Gln
805 810 815

Glu Leu Pro Pro Pro Gly Gln Gly Glu Gly Leu Leu Leu Gly Phe Phe
820 825 830

Gln Lys Ala Tyr Ala Pro Phe Leu Leu His Trp Ile Thr Arg Gly Val
835 840 845

Val Leu Leu Leu Phe Leu Ala Leu Phe Gly Val Ser Leu Tyr Ser Met
850 855 860

Cys His Ile Ser Val Gly Leu Asp Gln Glu Leu Ala Leu Pro Lys Asp
865 870 875 880

Ser Tyr Leu Leu Asp Tyr Phe Leu Phe Leu Asn Arg Tyr Phe Glu Val
885 890 895

Gly Ala Pro Val Tyr Phe Val Thr Thr Leu Gly Tyr Asn Phe Ser Ser
900 905 910

Glu Ala Gly Met Asn Ala Ile Cys Ser Ser Ala Gly Cys Asn Asn Phe
915 920 925

Ser Phe Thr Gln Lys Ile Gln Tyr Ala Thr Glu Phe Pro Glu Gln Ser
930 935 940

Tyr Leu Ala Ile Pro Ala Ser Ser Trp Val Asp Asp Phe Ile Asp Trp
945 950 955 960

Leu Thr Pro Ser Ser Cys Cys Arg Leu Tyr Ile Ser Gly Pro Asn Lys
965 970 975

Asp Lys Phe Cys Pro Ser Thr Val Asn Ser Leu Asn Cys Leu Lys Asn
980 985 990

Cys Met Ser Ile Thr Met Gly Ser Val Arg Pro Ser Val Glu Gln Phe
995 1000 1005

His Lys Tyr Leu Pro Trp Phe Leu Asn Asp Arg Pro Asn Ile Lys
1010 1015 1020

Cys Pro Lys Gly Gly Leu Ala Ala Tyr Ser Thr Ser Val Asn Leu
1025 1030 1035

Thr Ser Asp Gly Gln Val Leu Asp Thr Val Ala Ile Leu Ser Pro
1040 1045 1050

Arg Leu Glu Tyr Ser Gly Thr Ile Ser Ala His Cys Asn Leu Tyr
1055 1060 1065

Leu Leu Asp Ser Ala Ser Arg Phe Met Ala Tyr His Lys Pro Leu
1070 1075 1080

Lys Asn Ser Gln Asp Tyr Thr Glu Ala Leu Arg Ala Ala Arg Glu
1085 1090 1095

Leu Ala Ala Asn Ile Thr Ala Asp Leu Arg Lys Val Pro Gly Thr
1100 1105 1110

Asp Pro Ala Phe Glu Val Phe Pro Tyr Thr Ile Thr Asn Val Phe
1115 1120 1125

Tyr Glu Gln Tyr Leu Thr Ile Leu Pro Glu Gly Leu Phe Met Leu
1130 1135 1140

Ser Leu Cys Leu Val Pro Thr Phe Ala Val Ser Cys Leu Leu Leu
1145 1150 1155

Gly Leu Asp Leu Arg Ser Gly Leu Leu Asn Leu Leu Ser Ile Val
1160 1165 1170

Met Ile Leu Val Asp Thr Val Gly Phe Met Ala Leu Trp Gly Ile

1175 1180 1185

Ser Tyr Asn Ala Val Ser Leu Ile Asn Leu Val Ser Ala Val Gly
1190 1195 1200

Met Ser Val Glu Phe Val Ser His Ile Thr Arg Ser Phe Ala Ile
1205 1210 1215

Ser Thr Lys Pro Thr Trp Leu Glu Arg Ala Lys Glu Ala Thr Ile
1220 1225 1230

Ser Met Gly Ser Ala Val Phe Ala Gly Val Ala Met Thr Asn Leu
1235 1240 1245

Pro Gly Ile Leu Val Leu Gly Leu Ala Lys Ala Gln Leu Ile Gln
1250 1255 1260

Ile Phe Phe Phe Arg Leu Asn Leu Leu Ile Thr Leu Leu Gly Leu
1265 1270 1275

Leu His Gly Leu Val Phe Leu Pro Val Ile Leu Ser Tyr Val Gly
1280 1285 1290

Pro Asp Val Asn Pro Ala Leu Ala Leu Glu Gln Lys Arg Ala Glu
1295 1300 1305

Glu Ala Val Ala Ala Val Met Val Ala Ser Cys Pro Asn His Pro
1310 1315 1320

Ser Arg Val Ser Thr Ala Asp Asn Ile Tyr Val Asn His Ser Phe
1325 1330 1335

Glu Gly Ser Ile Lys Gly Ala Gly Ala Ile Ser Asn Phe Leu Pro
1340 1345 1350

Asn Asn Gly Arg Gln Phe
1355

<210> 23
<211> 21
<212> DNA
<213> artificial

<220>
<223> synthetic sequence

<400> 23
tggtctttac agaactcact a 21

<210> 24
<211> 21
<212> DNA
<213> artificial

<220>
<223> synthetic sequence

<400> 24
tccggacaat accagtctct a 21

<210> 25
<211> 76
<212> DNA
<213> artificial

<220>
<223> synthetic sequence

<400> 25
ggatcccgta gtgagttctg taaagaccat tgatatccgt ggtctttaca gaactcacta 60

ttttttccaa aagctt 76

<210> 26
<211> 76
<212> DNA
<213> artificial

<220>
<223> synthetic sequence

<400> 26
ggatcccgta gagactggta ttgtccggat tgatatccgt ccggacaata ccagtctcta 60

ttttttccaa aagctt 76

<210> 27
<211> 960
<212> DNA
<213> Homo sapiens

<400> 27
atctgcagct cagctttggt aatgggggcc cattaccaaa tgggggtaaa ggtcatggcc 60

catcctggtg atagtgagaa cccaaggtag gccttgaaga ttcctatcag gagggagcag 120

aaagtgtgta ccacacccct gggcccaggt ggagcagggc tgctgctcaa ggctcccagc 180

catgctctgt cccttgctag gggtgaccgg tgggacaggc ctgggcaagg gacaagaggg 240

agaaggtcgg ggggaagagg ggatgaagag caaagtgagc aaaggagagt cttccactat 300

ctggggtctc tgtcaactgt caggccctag agtgagctgt tctttccctt tgcttcctgg 360

aggaggggac ttttgtcact gcgtcactcc accctgcctg cccctccgtt atcaggctgt 420

taatattaat taacaacagt tgctagggat gacagtgcag agggttcctc tgagcccatt 480

gctggccctg gtcccaagag ggggtagggc agagctgggg tctgaggctg agccagggag 540

ggtgcggagg ttcctcggcc atgctgagct cctgaggccg ggtcccagcc agtgcctggt 600

cccatctgtg cctccaggcc ctggcaccaa ctccagcagt gttaggggct aatagcgtgg 660

tctctccccc agctgactca gccctctggc ttcggtcgct ttgggaagtg agtggagacc 720

ctagcacctg cgtgatgagg ctcatctaaa gcgggggcct gtggactggg gccaaacagt 780

gggagtggtg gatcattaac cagcagggct cagcctcatt ggtccctaac ccagtcaggc 840

cagggttgtc atcgaagggg aggaggctgc cttaatgtgt gttcagccct tggctgttcc 900

tgaggcctgg cctggctccc cgctgacccc ttcccagacc tgggatggcg gaggccggcc 960

<210> 28
<211> 970
<212> DNA
<213> Mus musculus

<400> 28
cctgcctaag cttgggcgga ttcccctctg agcccacccg agcccctggg acactggtgg 60

aactcagtag gagcccctcc ctgcagctgt ctcaacaggt agctgcatga gtggccttga 120

agcaattatc agcaattcag ccctggcaat agaggccaag gtcctggcct gtcttggtga 180

tagcaagagc ccaaggaaag actggaagtt tcctactgga aagaagcaga ggatgaacca 240

tgtacctggg cccaggttgg gtgggacttg ccactcagag cccctaacca gggttgttca 300

gaggactagg ccagggccag gaccaagaaa gggatagaac gggcatgagg aggaagggtg 360

aagggatcca aggaatctct ggtcctgttc cctgttagga catttgtcat ggaatcactc 420

tcgcttagtg tctctgttat ctgggtgcta atagcaacta ttcagttgct aggatgttag 480

gtgagtctga acctaccctt gatgttgatc tgaagaggcg atgcgttaga ctgcaggttg 540

gaggccaagt ccaggacagt gttgatattc tggatctcca agaagcctcc aaggccaaag 600

ccaggccagt gtctggtctc gcagaggaac agctctgcat ctcttgcccg gttggctcta 660

actaccacat tagacttcag ttgcgtcaaa aaacgagggg accccagcgc cttcactagg 720

aagttgacct cagaaggagg agatggaatg gcaccatctg atgtaaggga agagaaaata 780

aattattaac cagtacggcc cagtcctatt ggccccatga cagacgaggg ttatcactaa 840

gaggaggaag ctgccttaat gtgcaaactc aggggccagt cctcagcttc cccggctgtc 900

tccaaggcct ggtcctgctt ttccttgatc acttcctggc tctgggatgg cagctgcctg 960

gcagggatgg 970

<210> 29
<211> 8
<212> PRT
<213> artificial

<220>

<223> synthetic peptide tag

<400> 29

Asp Tyr Lys Asp Asp Asp Asp Lys
1 5

<210> 30
<211> 23
<212> DNA
<213> artificial

<220>
<223> primer

<400> 30
ctatacgaag ttatgtcaag cgg 23

<210> 31
<211> 25
<212> DNA
<213> artificial

<220>
<223> primer

<400> 31
cttgcacctg acttcctcat ataag 25

<210> 32
<211> 23
<212> DNA
<213> artificial

<220>
<223> primer

<400> 32
aaagaaggaa agcggccgcc agg 23

<210> 33
<211> 25
<212> DNA
<213> artificial

<220>
<223> primer

<400> 33
aggaaccgta ctgagcgcat accaa 25

INTERNATIONAL SEARCH REPORT

In:

PCT/US05/27579

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07K 14/00; C12N 5/10, 15/11, 15/693; G01N 33/53

US CL : 435/7.2, 69.1, 320.1, 325; 530/350; 536/23.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.2, 69.1, 320.1, 325; 530/350; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6,426,198 B1 (CARSTEA et al.) 30 July 2002 (30.07.2002) see entire document.	1-46

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 December 2005 (12.12.2005)	22 DEC 2005
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201	Authorized officer Michael Pak Telephone No. 571-272-1600

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US05/27979

Continuation of B. FIELDS SEARCHED Item 3:
WEST, PIR, UNIPROT, GENESEQ

search terms: Niemann-Pick C1-like protein, NPC1L1, lipid permease, metabolism, cholesterol